2019

DOKY: A Multi-Modal User Interface for Non-Visual Presentation, Navigation and Manipulation of Structured Documents on Mobile and Wearable Devices

Dorigo, Martin Lukas

http://hdl.handle.net/10026.1/13432

http://dx.doi.org/10.24382/809
University of Plymouth

UNIVERSITY OF PLYMOUTH

DOKY:
A Multi-Modal User Interface for Non-Visual Presentation, Navigation and Manipulation of Structured Documents on Mobile and Wearable Devices

by

MARTIN LUKAS DORIGO

A thesis submitted to the University of Plymouth
in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Computing, Electronics and Mathematics

December 2018

Copyright Statement

Abstract

DOKY: A Multi-Modal User Interface for Non-Visual Presentation, Navigation and Manipulation of Structured Documents on Mobile and Wearable Devices

Martin Lukas Dorigo

There are a large number of highly structured documents available on the Internet. The logical document structure is very important for the reader in order to efficiently handling the document content. In graphical user interfaces, each logical structure element is presented by a specific visualisation, a graphical icon. This representation allows visual readers to recognise the structure at a glance. Another advantage is that it enables direct navigation and manipulation. Blind and visually impaired persons are unable to use graphical user interfaces and for the emerging category of mobile and wearable devices, where there are only small visual displays available or no visual display at all, a non-visual alternative is required too.

A multi-modal user interface for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices like smart phones, smart watches or smart tablets has been developed as a result of inductive research among 205 blind and visually impaired participants. It enables the user to get a fast overview over the document structure and to efficiently skim and scan over the document content by identifying the type, level, position, length, relationship and content text of each element as well as to focus, select, activate, move, remove and insert structure elements or text. These interactions are presented in a non-visual way using Earcons, Tactons and synthetic speech utterances, serving the auditory and tactile human sense. Navigation and manipulation is provided by using the multitouch, motion (linear acceleration and rotation) or speech recognition input modality. It is a complete solution for reading, creating and editing structured documents in a non-visual way. There is no special hardware required. The name DOKY is derived from a short form of the terms document, and accessibility.

A flexible platform-independent and event-driven software architecture implementing the DOKY user interface as well as the automated structured observation research method employed for the investigation into the effectiveness of the proposed user interface has been presented. Because it is platform- and language-neutral, it can be used in a wide variety of platforms, environments and applications for mobile and wearable devices. Each component is defined by interfaces and abstract classes only, so that it can be easily changed or extended, and grouped in a semantically self-containing package.

An investigation into the effectiveness of the proposed DOKY user interface has been carried out to see whether the proposed user interface design concepts and user interaction design concepts are effective means for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices, by automated structured observations of 876 blind and visually impaired research subjects performing 19 exercises among a highly structured example document using the DOKY Structured Observation App on their own mobile or wearable device remotely over the Internet. The results showed that the proposed user interface design concepts for presentation and navigation and the user interaction design concepts for manipulation are effective and that their effectiveness depends on the input modality and hardware device employed as well as on the use of screen readers.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Especially I want to thank my director of studies, Prof. Dr. Bettina Harriehausen-Mühlbauer, for the support and the discussions which led to many new ideas. Your advice and guidance on both research as well as on my further way to go have been invaluable for me.

I would also like to thank my second supervisor, Prof. Dr. Ingo Stengel, for his constructive words, when I needed them and for his advice, guidance, and his valuable feedback on papers and the thesis.

Thanks must also go to my third supervisors, Dr. Ismini Vasileiou and Dr. Paul Steven Haskell-Dowland, who spent a lot of time proof reading papers and the thesis, in addition to providing helpful experience and guidance throughout my studies.

In addition, I would like to thank all the very supportive members of the Centre for Security, Communications and Network Research (CSCAN) and the administrative staff at the University of Plymouth. Many thanks to Ms. Carole Watson and Ms. Lucy Cheetham for their support by helping me with administrative issues.

My thanks go also to Prof. Dr. Alireza Darvishy and Prof. Dr. Hans-Peter Hutter from Zurich University of Applied Sciences (ZHAW) as well as to Prof. Dr. Brigitte Tanner for establishing the connection to ZHAW. Thank you for being my mentor and that I had the chance to learn so much from you.

Moreover, I would like to thank Prof. Dr. Andreas Rinkel from University of Applied Sciences Rapperswil (HSR) for establishing the connection to the University of Plymouth. Without you, I would not have even started this long, exciting journey.

Thanks also goes to my fellow students from the Ph.D. seminar at University of Applied Sciences Darmstadt, Germany for the discussions which led to many new ideas.

I wish to thank the 876 blind and visually impaired research subjects in the United Kingdom, the United States, Germany, Austria and Switzerland as well as in other regions of the world, forming the sample for this research, for participating in this research by answering the questionnaire for the initial survey and automated structured observations by performing 19 exercises among a highly structured example document using the DOKY Structured Observation App on their own mobile or wearable device remotely over the Internet.

I also wish to thank all the different main organisations in support of blind and visually impaired persons in the United Kingdom, the United States, Germany, Austria and Switzerland as well as their sub-organisations in each county for their cooperation in sending out the invitation for this research to all of their members by distributing the invitation letter across their mailing lists, news letters, forums, magazines and web-sites.

Many thanks to Mr. François Küssenberger from François Küssenberger & Partner Live Marketing & Communication GmbH for proposing invitation letter texts for inviting participants and organisations which leaded to high response rates.

Also, I would like to thank Dr. Benjamin George Sanders, who cared about me and supported me during my stay in Plymouth in the United Kingdom.

Finally, last but not least, I wish to thank my friends and family for their support.

Without all of you, this research would not have been possible. Thank you so much for your valuable support.

Authors Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Doctoral College Quality Sub-Committee.

Work submitted for this research degree at the University of Plymouth has not formed part of any other degree either at the University of Plymouth or at another establishment.

Relevant scientific seminars and conferences were regularly attended at which work was often presented and several papers prepared for publication.

Bibliographic details of publications and presentations carried out during the research programme can be found below.

Word count of main body of thesis: 79'393

Signed: Martin Dorigo

Date: December 21st 2018

Publications

The following papers relating to the research programme have been published and presented at refereed conferences:

- Dorigo, M.L., Harriehausen-Mühlbauer, B., Stengel, I., Dowland, P. S. 2015. Evaluation of Doky: A Multi-Modal User Interface for Non-Visual Presentation, Navigation and Manipulation of Structured Documents on Mobile and Wearable Devices. In Proceedings of the 17th International Conference on Human-Computer Interaction (Los Angeles, California, USA, July 02 – 07, 2015). HCII 2015.

- Dorigo, M.L., Harriehausen-Mühlbauer, B., Stengel, I., Dowland, P. S. 2014. Nonvisual Presentation, Navigation and Manipulation of Structured Documents on Mobile and Wearable Devices. In Proceedings of the 14th International Conference on Computers Helping People with Special Needs (Saint-Denis, Paris, France, July 09 – 11, 2014). ICCHP 2014. https://link.springer.com/chapter/10.1007/978-3-319-08596-8_59

- Dorigo, M.L., Harriehausen-Mühlbauer, B., Stengel, I., Dowland, P. S. 2013. Nonvisual Presentation and Navigation within the Structure of Digital Text-Documents on Mobile Devices. In Proceedings of the 15th International Conference on Human-Computer Interaction (Las Vegas, Nevada, USA, July 21 – 26, 2013). HCII 2013. https://link.springer.com/chapter/10.1007/978-3-642-39194-1_37

- Dorigo, M.L., Harriehausen-Mühlbauer, B., Stengel, I., Dowland, P.S. 2012. Nonvisual Presentation and Navigation within the Structure of Digital Text-Documents on Mobile Devices. In Proceedings of the 10th Postgraduate Society Conference (Plymouth, Devon, United Kingdom, November 21st, 2012).

- Dorigo, M.L., Harriehausen-Mühlbauer, B., Stengel, I., Dowland, P. S. 2012. Nichtvisuelle Darstellung und Navigation der Struktur von Elektronischen Text-Dokumenten auf Mobilen Endgeräten. Tag der Forschung 2012 (Darmstadt, Hessen, Germany, November 20th, 2012).

- Dorigo, M., Harriehausen-Mühlbauer, B., Stengel, I., Dowland, P. S. 2012. Nonvisual skimming and scanning over the structure of digital text-documents on mobile devices. Young Researcher Consortium of the 13th International Conference on Computers Helping People with Special Needs (Linz, Austria, July 09th - 13th, 2012). ICCHP 2012.

- Dorigo, M., Harriehausen-Mühlbauer, B., Stengel, I., Dowland, P. S. 2011. Non-visual skimming and scanning over the structure of digital text-documents. Doctoral Consortium (DC) of the 13th International ACM SIGACCESS Conference on Computers and Accessibility (Dundee, Scotland, United Kingdom, October 24 – 26, 2011). ASSETS 2011.

- Dorigo, M., Harriehausen-Mühlbauer, B., Stengel, I., Dowland, P. S. 2011. Survey: Improving document accessibility from the blind and visually impaired user's point of view. In Proceedings of the 14th International Conference on Human-Computer Interaction (Orlando, Florida, USA, July 09 – 14, 2011). HCII 2011. https://link.springer.com/chapter/10.1007%2F978-3-642-21657-2_14

- Dorigo, M., Harriehausen-Mühlbauer, B., Stengel, I., Dowland, P. S. 2011. Survey: Assistive Technologies for blind and visually impaired people. In Proceedings of the 26th California State University Northridge Center on Disabilities Annual International Technology & Persons with Disabilities Conference (San Diego, California, USA, March 14 – 19, 2011). CSUN 2011. http://www.letsgoexpo.com/utilities/File/viewfile.cfm?LCID=4756

- Davishy, A., Hutter, H.P., Horvath, A., Dorigo, M.: A Flexible Software Architecture Concept for the Creation of Accessible PDF Documents. In proceedings of the 12th International Conference on Computers Helping People with Special Needs (Wien, Austria, July 14 – 16, 2010), LNCS 6179, Springer Verlag Berlin. ICCHP 2010. https://link.springer.com/chapter/10.1007/978-3-642-14097-6_8

1 Introduction

There are a large number of highly structured documents such as text-documents, spreadsheets, presentations, graphics, mathematical formulae or equations and web-sites available on the Internet. The logical document structure like headings, paragraphs, lists or tables is very important for the reader in order to efficiently handling the document. If the reader has an overview over the logical structure of the entire document in mind, he is able to efficiently skim and scan over the document and to quickly find relevant information within the document content, he or she is searching for.

In graphical user interfaces used on many modern computers, each logical document structure element is presented by a specific visualisation, a so called graphical icon (Shneiderman, 1998). This representation allows visual readers to recognise the structure at a glance and much faster than if the structure is described in text only. Another advantage of a graphical user interface is that it enables direct navigation and manipulation of the document structure and content. These visualisations can look like the following examples shown in Figure 1.1:

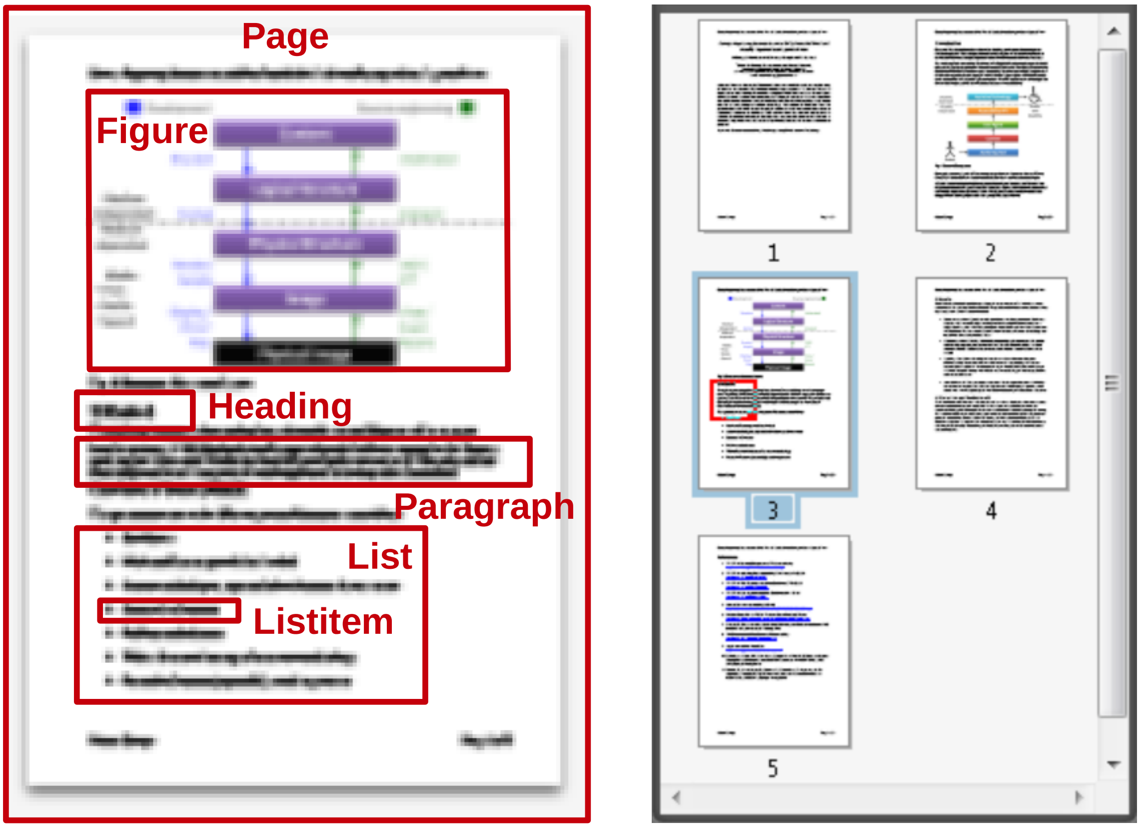


Figure 1.1: Visual Presentation and Overview over the Logical Document Structure

A heading is presented by using a bigger font on a separate line, where the font size indicates the level of the heading within the hierarchy. A paragraph of text is presented by a number of coherent lines, starting and ending with a space above and below the paragraph. A list is composed of multiple list items, each starting on a separate line. Each list item is visualised at an indented position and leaded by a bullet symbol or a number.

Especially for large and highly structured documents like newspaper articles or scientific literature, the logical structure can be quite complex. Therefore most user agents for the reading of structured documents provide a visual overview function, where the logical structure of the document only, without the document content, is visually presented at a higher level of abstraction. The content (text and graphics) is substituted by place holders like dots, lines or thumbnails, indicating the length and size of it. This feature allows visual readers to get a fast overview over the logical structure of the entire document at a glance, skim and scan over the document content, see the current position of the cursor and to quickly navigate within the document.

Blind and visually impaired persons, on the other hand, are unable to use a graphical user interface because it is inaccessible to them and cannot see the visual presentation of the logical document structure. The lack of this very important information prevents them from finding relevant information fast and causes frustration (Lazar, Allen, Kleinman & Malarkey, 2007). Providing information in an auditory and tactile form could generally help solve this problem and allow blind and visually impaired persons to use the same facilities as persons without disability, what is known as inclusive design.

For the emerging category of mobile and wearable devices, where there are only small visual displays available or no visual display at all, a non-visual alternative for presentation, navigation and manipulation of structured documents is required too. It has the potential to improve interaction in a range of different areas, particularly where the visual display is overloaded, limited in size or not available, such as in user interfaces for blind and visually impaired persons or on mobile and wearable devices.

There is also psychological evidence showing that there are advantages to be gained from using combined auditory and tactile multi-modal user interfaces and that work in their development is justified. Psychological evidence of Brown, Newsome & Glinert (1989), Perrott, Sadralobadi, Saberi & Strybel (1991) and Brewster (1992) suggest that sharing information across different sensory modalities can actually improve task performance. Having redundant information gives the user multiple chances of identifying the data. For example, if they cannot remember how a document structure element sounds like they may be able to remember what it feels like.

Some early work was done by Colquhoun (1975) in which simple sounds were added to a visual sonar monitoring system. Users had to monitor either an auditory, a visual or a dual-mode display and indicate the presence of a target stimulus. The test results showed that the audio-visual display got the highest target detection rates, although error rates were similar. Colquhoun (1975) suggested that dual-mode displays with redundant signals (signals simultaneously presented on auditory and visual displays) have been found superior either to both single-mode displays or to at least one of them. In no case has the dual-mode display been found significantly inferior to a single-mode display. This research indicates that there are definite advantages to be gained from using multi-modal user interfaces.

Brown, Newsome & Glinert (1989) performed visual search experiments using auditory or visual target cues. Their aim was to reduce visual workload by using multiple sensory modalities as suggested by the multiple resource theory (Wickens, Mountford & Schreiner, 1981). They define the theory thus: Humans possess different capacities, each with separate resource properties. If tasks demand separate resources, performance of two simultaneous tasks will be more efficient. Task interference occurs when the same resources are called upon simultaneously. They suggest that inter-modal task sharing (dividing attention to a task between eyes and ears) is more successful than intra-modal (the eyes doing two things at once). The experiments they conducted showed that the auditory modality could, in some cases, be more effective than the visual one. Nonetheless, their findings do suggest that humans can extract more than one piece of information from a sound and then act upon it.

The use of non-visual information cues at the user interface is becoming increasingly popular due to the potential benefits it offers. It can be used to present information otherwise unavailable on a visual display for example mode information (Monk, 1986) or information that is hard to discern visually, such as multi-dimensional numerical data (Frysinger, 1990). It is a useful complement to visual output because it can increase the amount of information communicated to the user or reduce the amount the user has to receive through the visual channel. It makes use of the auditory and the tactile system which is powerful but under-utilised in most current user interfaces.

The foveal area of the retina (the part of greatest acuity) subtends an angle of only two degrees around the point of fixation (Rayner & Pollatsek, 1989). Sound, on the other hand, can be heard from 360 degrees without the need to concentrate on an output device, thus providing greater flexibility. Sound is also good at capturing a user's attention whilst they are performing another task.

In this thesis, the term structured document is used in a broad sense. Is is being used as a way of representing many different kinds of information that may be stored in diverse systems, and much of this would traditionally be seen as structured data rather than as documents but the DOKY user interface may be used to manage this data.

1.1 Aims and Objectives

The aim of this research project is to develop a multi-modal user interface for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices like smart phones, smart watches or smart tablets. It should enable the user to get a fast overview over the document structure and to efficiently skim and scan over the document content by identifying the type, level, position, length, relationship and content text of each element as well as to focus, select, activate, move, remove and insert structure elements or text. These interactions should be presented in a non-visual way serving the auditory and tactile human sense. Navigation and manipulation should be provided by using the multitouch, motion or speech recognition input modality. It is a complete solution for reading, creating and editing structured documents in a non-visual way. There should be no special hardware required.

The potential is not limited to blind and visually impaired persons. It also benefits persons without disability in situations where the visual sense is in use by other tasks as well as on the emerging category of mobile and wearable devices, where there are only small visual displays available or no visual display at all. In order to achieve this aim, the research can be divided into the following five distinct objectives:

1. Investigation into what structured documents are in general and identification of the essential components involved in the process of non-visual reading, creating and editing of structured documents. Examination of the different approaches available for each component in greater detail and identification of their potential applications in a user interface for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices as well as some examples of existing systems that use each non-visual approach in their human computer user interfaces.
2. Investigation into how blind and visually impaired persons handle structured documents at the moment and what is of importance as to the reading of and the navigation within structured documents from the blind and visually impaired person's point of view in order to develop a novel concept for a multi-modal user interface for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices.

3. Development of a multi-modal user interface for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices like smart phones, smart watches or smart tablets as a result of inductive research among blind and visually impaired participants.
4. Proposal of a flexible platform-independent and event-driven software architecture implementing the user interface as well as an automated structured observation research method employed for the investigation into the effectiveness of the proposed user interface. It should be platform- and language-neutral, so that it can be used in a wide variety of platforms, environments and applications for mobile and wearable devices. Each component should be defined by interfaces and abstract classes only, so that it can be easily changed or extended, and grouped in a semantically self-containing package.
5. Investigation into the effectiveness of the proposed user interface to see whether the proposed user interface design concepts and user interaction design concepts are effective means for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices, by automated structured observations of blind and visually impaired research subjects performing exercises among a highly structured example document using the DOKY Structured Observation App on their own mobile or wearable device remotely over the Internet.

1.2 Thesis Structure

The thesis addresses the aforementioned aims and objectives in order and is consisted of the following chapters:

- Chapter 2 starts with an introduction about what structured documents are in general and the essential components involved in the process of non-visual reading, creating and editing of structured documents are identified. The different approaches available for each component are described in greater detail and their potential applications in a user interface for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices are discussed. In addition, some examples of existing systems that use each non-visual approach in their human computer user interfaces are provided.
- Chapter 3 proposes a research methodology for developing the user interface by employing a natural science epistemological model (positivism), an objective ontology (objectivism), a quantitative research strategy and an iterative research approach. The user interface has been created using an inductive research approach where general research questions stand at the beginning and the theory, in this case the user interface design and the user interaction design, is the outcome of the research. For the evaluation of the user interface, a deductive research approach has been followed where the purpose of the theory is to generate the hypotheses that can be tested and that will thereby allow explanations of laws to be assessed. The last step involved a movement that is in the opposite direction from deduction to induction, as the implications of the findings from the theory that prompted the whole exercise had been inferred. These findings were fed back into the stock of theory and the research findings associated with the domain of non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices.
- Chapter 4 presents the results of a a survey, which has been conducted among 205 blind and visually impaired persons to find out how they handle structured documents at the moment and what is of importance as to the reading of and the navigation within structured documents from the blind and visually impaired person's point of view, in order to develop a novel concept for a multi-modal user interface for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices.
- Chapter 5 proposes a multi-modal user interface for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices like smart phones, smart watches or smart tablets which has been developed as a result of inductive research among 205 blind and visually impaired participants in the previous chapters. It enables the user to get a fast overview over the document structure and to efficiently skim and scan over the document content by identifying the type, level, position, length, relationship and content text of each element as well as to focus, select, activate, move, remove and insert structure elements or text. These interactions are presented in a non-visual way using Earcons, Tactons and synthetic speech utterances, serving the auditory and tactile human sense. Navigation and manipulation is provided by using the multitouch, motion (linear acceleration and rotation) or speech recognition input modality. It is a complete solution for reading, creating and editing structured documents in a non-visual way. There is no special hardware required.

- Chapter 6 presents the results of an investigation into the effectiveness of the proposed DOKY user interface, which was carried out to see whether the proposed user interface design concepts and user interaction design concepts are effective means for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices, by automated structured observations of 876 blind and visually impaired research subjects performing 19 exercises among a highly structured example document using the DOKY Structured Observation App on their own mobile or wearable device remotely over the Internet.
- Chapter 7 concludes the thesis by summarising the achievements of the research programme and discussing the limitations of the performed research. In addition, considerations about possible directions of further research based on the presented research results within the area of non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices are proposed.
- The questionnaire used for the initial survey among blind and visually impaired persons, presented in Chapter 4, is attached in Appendix A.
- The observation schedule used for the automated structured observation of blind and visually impaired research subjects, presented in Chapter 6, is attached at the end of this thesis in Appendix B.
- Appendix C presents a flexible platform-independent and event-driven software architecture implementing the DOKY user interface as well as the automated structured observation research method employed for the investigation into the effectiveness of the proposed user interface. Because it is platform- and language-neutral, it can be used in a wide variety of platforms, environments and applications for mobile and wearable devices. Each component is defined by interfaces and abstract classes only, so that it can be easily changed or extended, and grouped in a semantically self-containing package.
- In addition, a number of papers relating to the research programme have been published and presented at refereed conferences. A list and copies of these publications can be found in Appendix D.

2 Literature Review: State of the Art of Non-Visual Access to Structured Documents

2.1 Introduction

This chapter starts with an introduction about what structured documents are in general and the essential components involved in the process of non-visual reading, creating and editing of structured documents are identified. The different approaches available for each component are described in greater detail and their potential applications in a user interface for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices are discussed. In addition, some examples of existing systems that use each non-visual approach in their human computer user interfaces are provided.

2.1.1 Essential Components

On its way from the author to the reader, a digital structured document passes several components. According to Henry & Duffy (2005), these components have to work together to provide full access for the user. If an accessibility feature is not implemented in one component, it does not result in an accessible user experience. Figure 2.1 gives an overview over the essential components involved in the process of non-visual presentation, navigation and manipulation of structured documents.

There are many different content types of digital structured documents available on the Internet, for example text-documents, spreadsheets, presentations, drawings, mathematics or web-sites and manifold different formats are used to store this content and exchanging it over the Internet.

The user agent or authoring tool parses this content to read a structured document or transforms the structured document to store it in a specific content format and enables assistive technologies to dynamically access and update the structure, content and style of the document by proving the relevant information and control via the accessibility API of the operating system. The research in this thesis is not about user agents and authoring tools because the main focus is on developing an assistive technology.

The accessibility Application Programming Interface (API) of the operating system, acts as the global interface between user agents, authoring tools and the assistive technologies running on a system. It enables assistive technologies to dynamically access and update the structure, content and style of a structured document by proving the relevant information and control.

The assistive technology, like screen reader, screen magnifier or the DOKY user interface developed in this research, dynamically accesses and updates the structure, content and style of the structured document provided by the accessibility API of the operating system and provides an alternative user interface for presentation, navigation and manipulation to the user by employing various different modalities.

The modality, for example graphical text, synthetic speech or Braille, presents particular pieces of information to the user in a modality specific way by stimulation of one or more human senses using specific hardware output devices.

The reader (a person who reads a structured document) or the author (a person who creates and edits structured documents) knows how to use the assistive technologies, user agents or authoring tools as well as the different modalities employed for presentation, navigation and manipulation to read, create and edit structured documents.

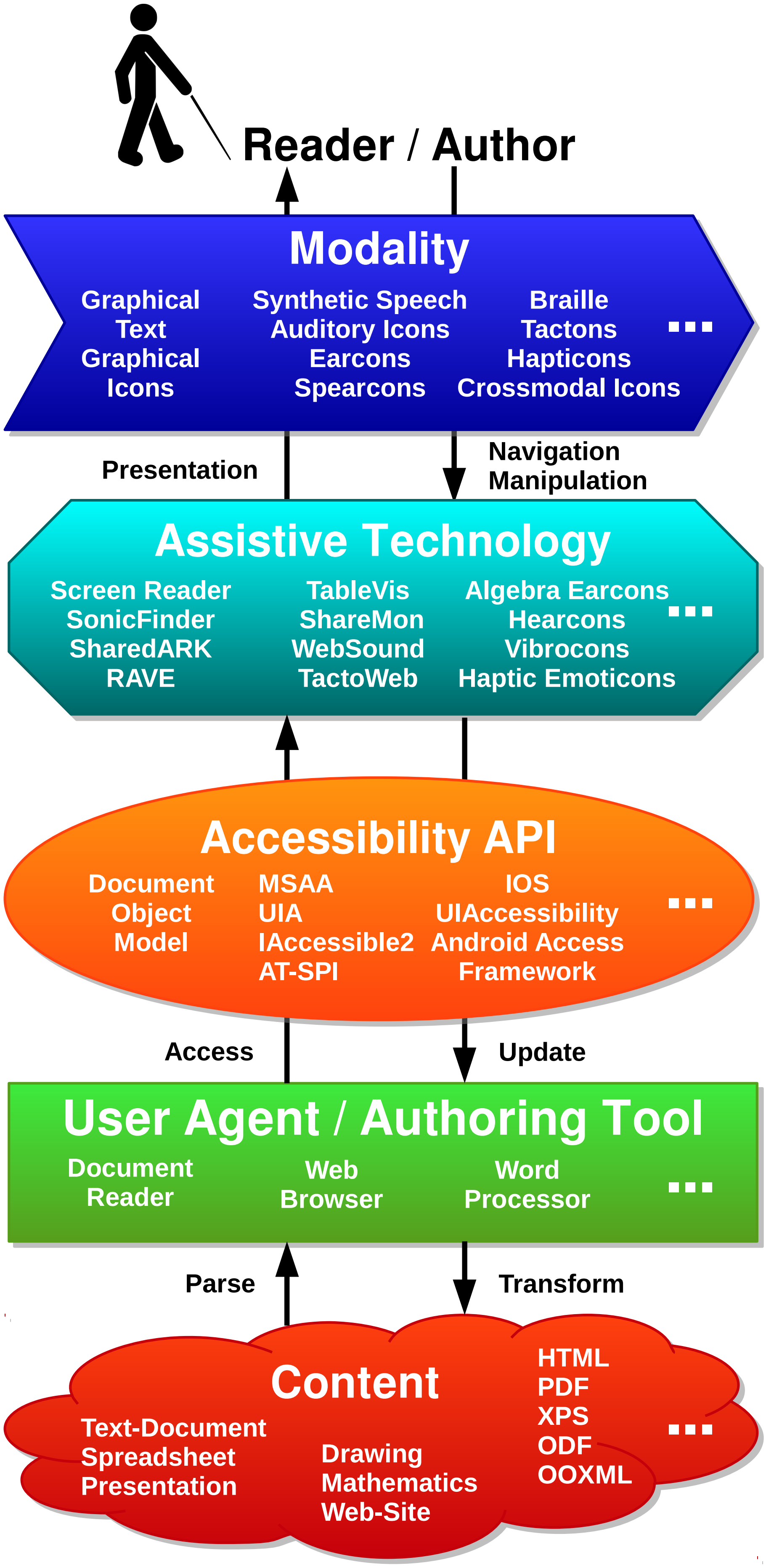


Figure 2.1: Essential Components for Non-Visual Access to Structured Documents

2.1.2 The Structured Document

During the processing by the different components for document development and document recognition, structured documents take on different forms. According to Dori, David, Doermann, Shin, Haralick, Phillips, Mitchell, Buchmann & Ross (1997) and Brugger (1998), a static digital structured document can be broken down into the following layers of abstraction. Figure 2.2 gives an overview over the different abstraction layers, a static digital structured document takes on during the process of document development and document recognition by the different components:

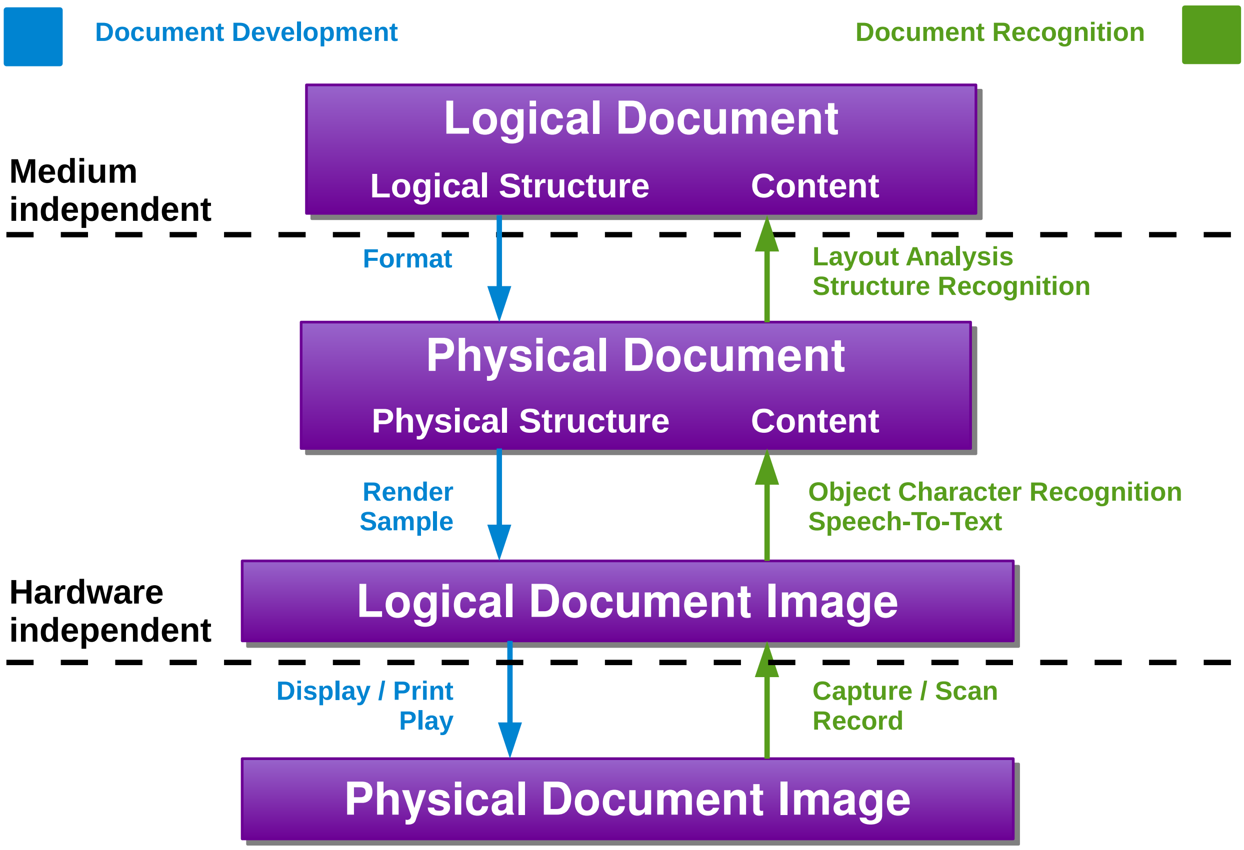


Figure 2.2: Abstraction Layers of a Structured Document

Logical Document

The logical document is formed by the logical document structure together with the document content. The document content describes the information contained within a structured document in a medium and presentation independent way. Elements of the document content are text and graphics as well as text alternatives for graphical content. The logical document structure defines how the content of a structured document is semantically organised in a medium and presentation independent way. Elements of the logical document structure are for example: Sections, headings, paragraphs, lists and tables. The logical document contains the most semantic information and is therefore the most abstract form. Hence, the logical document structure as well as the document content is directly accessible and can be used by assistive technologies for presenting it to the user employing different modalities.

Physical Document

The physical document is formed by the physical document structure together with the document content. The physical structure of a document specifies how the document content is physically organised on a specific medium. Elements of the physical document structure are for example: Pages, columns, blocks, lines, words and characters. The physical document does not semantically contain information about the logical structure. Hence, the logical document structure is not directly accessible.

Logical Document Image

The digital logical document image is the rendered or sampled form of the physical document in a hardware independent way. The elements of the logical document image depend on the type of medium employed. On visual media, elements of the logical document image are pixels with different colours (in most cases black or white), in the auditory domain, these are samples with different intensity and in the haptic domain, these may be dots with different heights and sizes. The logical document image does not contain any semantic information about the document, neither about the logical or physical structure nor about the content. Hence, the document is not directly accessible.

Physical Document Image

The analogue physical document image is created by presenting the digital logical document image to the user on a specific hardware output device. The elements of the physical document image depend on the type of hardware device employed. In the visual domain, elements of the physical image are led or mono crystal pixels and on screen, or using a printer, these are ink or toner dots on paper. In the auditory domain, using speakers or headphones, elements of the physical image are auditory waves according to the digital samples. In the tactile domain, using a dynamic Braille display, elements of the physical image are electromagnetic coil and pin dots, with an embosser, these are dots and holes in the paper and, using a vibrotactile transducer, these are mechanical vibrations on the skin.

Document Development

The process of developing a structured document and presenting it to the user on a specific medium and hardware device is known as document development. It is done by an author creating a structured document as well as automatically by a user agent, authoring tool or assistive technology which presents the document to the user. It consists of the following steps: The author writes the content and organises it in a logical structure. This logical document is than formatted into a physical document on a specific medium. Afterwards, the physical document is rendered or sampled to obtain the logical document image. This digital logical document image is than presented to the user in form of a physical document image on an specific analogue physical hardware device. Document development can easily by done automatically by a machine.

Document Recognition

The opposite procedure of the document development is named as document recognition. It is done by a person who reads a structured document (reader) as well as automatically by document recognition software, for example Nuance [OmniPage]. The following steps are required: The analogue physical document image is scanned, captured or recorded to obtain a digital logical document image. Object character recognition (OCR) or speech-to-text (STT) is performed to recognise the physical document out of this logical document image. To get the logical document out of the physical document, layout analysis and structure recognition is preformed. Finally, the logical document is read and understood by the person who reads the structured document (reader). While the document development can easily be done automatically by a machine, the document recognition on the other hand is very computationally intensive and error prone.

2.2 Modalities

The modality, presents particular pieces of information to the user in a modality specific way by stimulation of one or more human senses using specific hardware output devices. One major question that must be answered when designing a user interface is what modalities should be used at the user interface to present the information effectively. The different approaches available can be grouped into two main categories: Text-based and iconic modalities.

Text-based modalities are a good choice for presenting text-data, like the text content of document elements or the text alternative of graphical content. On the other hand, they may not be effective in presenting non-text data like the element type, level, position, length or relationship because Barker & Manji (1989) claim that an important limitation of text is its lack of expressive capability: It may take many words to describe a fairly simple concept.

Therefore iconic displays were introduced presenting data in a non-verbal way that speeded up interactions in a user interface as an icon is presented to the user of the thing they are interested in. The user then recalls its meaning rather than having the meaning described in text. An icon is also universal: It means the same thing in different languages.

Table 2.1 shows the different text-based and iconic approaches available for presenting data serving the visual, auditory and haptic human sense and abstract approaches:

Human Sense	Text-Based	Iconic
Visual	• **Graphical Text**	• **Graphical Icons** (Shneiderman, 1998)
Auditory	• **Synthetic Speech**	• **Auditory Icons** (Gaver, 1986) • **Earcons** (Blattner, Sumikawa & Greenberg, 1989) • **Spearcons** (Walker, Nance & Lindsay, 2006)
Haptic	• **Braille** (Foulke, 1982) • **Vibratese** (Geldard, 1957)	• **Tactons** (Brewster & Brown, 2004) • **Hapticons** (MacLean & Enriquez, 2003)
(Abstract)	• **Logical Text**	• **Crossmodal Icons** (Hoggan & Brewster, 2006)

Table 2.1: Text-Based and Iconic Output Modalities for Presenting Information

In the visual domain there is graphical text and its iconic counterparts are graphical icons. Shneiderman (1998) defines a graphical icon as an image, picture or symbol representing a concept. However, this research is not about visual modalities. In the auditory domain there is synthetic speech as text-based modality and there are auditory icons, earcons and spearcons as iconic counterparts. In the haptic domain one common form of text-based output is Braille in addition to Vibratese and the iconic counterparts are tactons and hapticons. The concept of crossmodal icons had been introduced that are abstract icons which can be automatically instantiated as either an earcon or tacton.

Another important point to bear in mind when designing a user interface is that the feedback to be used at the human computer interface must be able to keep pace with the interactions that occur. If it does not, either the system has to wait for the feedback to finish before continuing, or the feedback playing refers to an interaction that has already completed, then it will not be effective. It will not provide the user with any advantage so there will be no reason for using it. Interactions tend to happen quickly. Slowing interactions down so that the feedbacks could keep up would be unacceptable. Therefore the feedbacks used must be able to keep pace with them.

In this chapter the different approaches for presenting data in a non-visual way are described and the potential uses in a user interface for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices are discussed. In addition, some examples of existing systems that use each non-visual approach in their human computer user interfaces are provided.

2.2.1 Auditory Approaches

In the auditory domain there is synthetic speech as text-based modality and there are auditory icons, earcons and spearcons as iconic counterparts.

2.2.1.1 Auditory Icons

Gaver (1986) has developed the idea of auditory icons. These are natural, everyday sounds which are used to represent actions and objects within an interface. Auditory icons use environmental sounds that have a semantic link with the object they represent. Although no formal analysis has been undertaken, auditory icons have been shown to be an effective form of presenting information in sound in several different applications and environments.

Mountford & Gaver (1990) suggest that sound can provide information about many different things within the environment. For example physical events: Whether a dropped glass broke or bounced. Invisible structures: Tapping on a wall to find out if there is a hollow behind it. Dynamic changes: As a glass is filled with liquid a listener can hear when it is full. Abnormal structures: A malfunctioning engine sounds different to a normal one. Events in space: The sound of footsteps indicate that someone is approaching.

Gaver uses sounds of events that are recorded from the natural environment, for example tapping or smashing sounds. His research is based on the ideas of Vanderveer (1979) and Warren & Verbrugge (1984). This work suggests that people do not listen to the pitch and timbre of sounds but to the sources that created them. When pouring liquid a listener hears the fullness of the receptacle, not the increases in pitch. Warren & Verbrugge (1984) suggest that identification of sound sources, and the behaviour of those sources, is the primary task of the auditory system.

Another important property of everyday sounds is that they can convey multidimensional data. When a door slams a listener may hear: The size and material of the door, the force that was used and the size of room on which it was slammed. This could be used within a user interface so that selection of an object makes a tapping sound, the type of material could represent the type of object and the size of the tapped object could represent the size of the object within the interface. Björk (1985) and Ballas & Howard (1987) have also done work in this area.

Gaver (1992) and (1993) was working on a system of parametrising auditory icons. He says that auditory icons not only reflect categories of events and objects as visual icons do, but are parametrised to reflect their relevant dimensions as well. That means, if a file is large, it sounds large. If it is dragged over a new surface, we hear that new surface.

And if an ongoing process starts running more quickly, it sounds quicker. If auditory icons are to be generated in real-time, rather than just be stored as fixed sound samples, then ways must be found of controlling their parameters and synthesising them directly. This would allow families of auditory icons to be created which varied along certain attributes. Gaver puts forward some early work to solve this problem. He has developed algorithms to allow the description of sound properties. For impact sounds he can define the hardness of the hammer hitting an object, the material the object is made from and its size. He can define whether an object sounds like it is bouncing, breaking or spilling and also scraping and dragging. He has used frequency modulation synthesis techniques (Chowning, 1975) to simulate machine sounds and can control the size of the machine and its speed of operation. This work is in its early stages but if it continues may allow auditory icons to become very much more flexible than they are now.

Auditory icons are mostly effective at conveying content, rather than hierarchical structures. In a user interface for non-visual presentation, navigation and manipulation of structured documents, there are problems with the intuitive mappings. For example, there are no natural real world sounds to match different element types like sections, headings, paragraphs, lists and tables or different operations like focus, select, activate, move, remove and insert. Hence, this approach cannot be used in this research.

2.2.1.2 Earcons

Earcons were developed by Blattner, Sumikawa & Greenberg (1989), Sumikawa, Blattner, Joy & Greenberg (1986) and Sumikawa (1985). They use abstract, synthetic tones in structured combinations to create auditory messages. Blattner et al. define earcons as non-verbal audio messages that are used in the computer user interface to provide information to the user about some computer object, operation or interaction. Unlike Gaver's auditory icons, introduced in the previous chapter, there is no intuitive link between the sound and what it represents. The link must be learned by the listener. Earcons use a more musical approach than auditory icons.

Earcons are constructed from simple building blocks called motives. These are short, rhythmic sequences of pitches with variable intensity, timbre and register that can be combined in different ways. Blattner, Sumikawa & Greenberg (1989) describe a motive thus as a brief succession of pitches arranged in such a way as to produce a tonal pattern sufficiently distinct to allow it to function as an individual recognisable entity. The most important features of motives are rhythm, pitch, timbre, register and dynamics. Sumikawa (1985) defines rhythm and pitch as the fixed parameters of motives and timbre, register and dynamics as the variable parameters. The fixed parameters are what define a motive, the variable parameters change it.

Brewster, Wright & Edwards (1992) and (1993) conducted two experiments to investigate the effectiveness of earcons for presenting information in sound. The first experiment tried to answer if earcons are a good method of communicating complex information in sound, if musical timbres are more effective than simple tones and if rhythm is important in the recognition of earcons. The second experiment used the results of the first to create new earcons to overcome some of the difficulties that came to light.

From the results of the two experiments and studies of literature on psychoacoustics, some guidelines have been drawn up for use to help designers when creating earcons. These should be used along with the more general guidelines for the syntactic design and integration of audio cues in computer user interfaces given by Sumikawa, Blattner, Joy & Greenberg (1986) and Sumikawa (1985). A designer could use the guidelines to create earcons that could effectively communicate complex information in sound. They allow a designer with no knowledge of sound design to create a set of earcons that will be effective. He or she can be sure that the earcons will be perceivable and recognisable by listeners because they incorporate knowledge of earcon design and psychoacoustics.

One overall result which came out of the work is that much larger differences than those suggested by Blattner, Sumikawa & Greenberg (1989) must be used to ensure recognition. If there are only small, subtle changes between earcons then they are unlikely to be noticed by anyone but skilled musicians (if absolute judgements must be made).

Timbre

Motives can be made to sound different by the use of different timbres, for example playing one motive with the sound of a violin and the other with the sound of a piano. Blattner, Sumikawa & Greenberg (1989) suggest the use of simple timbres such as sine, square, sawtooth and triangular wave but psychoacoustics (the study of the perception of sound) suggests that complex musical instrument timbres may be more effective and easily recognised (Moore, 1989) than the simple tones proposed by Blattner. This is due to their greater spectral and temporal complexity making them more discriminable than simple tones. Brewster, Wright & Edwards (1993) also proposes to use complex musical instrument timbres. Where possible timbres with multiple harmonics should be used as this helps perception and can avoid masking. Timbres should be used that are subjectively easy to tell apart. For example, on a musical instrument synthesiser brass and organ rather than brass1 and brass2. However, instruments that sound different in real life may not when played on a synthesiser, so care should be taken when choosing timbres. Using multiple timbres per earcon may confer advantages when using compound earcons. McGookin (2004) found that users were able to achieve absolute identification rates of over 90 per cent for timbre.

Pitch

There are 96 different pitches in the western musical system and these could be combined to produce a large number of different motives. Sumikawa (1985) suggest that notes should be kept within a range of eight octaves of twelve notes. Sumikawa, Blattner, Joy & Greenberg (1986) also suggest that random combinations of pitches should not be used, but that they should be taken from one octave and in the same scale for easier manipulation. This fits in well with the work of Dewar, Cuddy & Mewhort (1977) who suggested that listeners can better detect differences between groups of sounds if all the notes in one sound are in a different scale to the other. Keeping all the notes in one octave also minimises pitch perception problems where a listener can mistake the octave to which a note belongs (Deutsch, 1986). Semitone gaps should be avoided as they can create incorrect melodic implications. Earcons should be musically neutral. Brewster, Wright & Edwards (1993) suggests that pitch should not be used on its own unless there are large differences between those used (see register below). Complex intra-earcon pitch structures are effective in differentiating earcons if used along with rhythm. Some suggested ranges for pitch are maximum 5 kHz (four octaves above middle C) and minimum 125 Hz - 150 Hz (the octave of middle C).

Register

The register is the position of the motive in the musical scale. A high register means high pitched notes and a low register low pitched notes. The same motive in a different register can convey a different meaning. Sumikawa (1985) suggest the only three registers (low, medium and high) should be used. Brewster, Wright & Edwards (1993) proposes that if listeners must be able to make absolute rather than relative judgements of earcons then pitch and register should not be used. A combination of pitch and another parameter would give better performance. If register alone must be used then there should be large differences between earcons but even then it might not be the most effective method. Two or three octaves difference should be used. This is not a problem if relative judgements are to be made. McGookin (2004) found that absolute identification rates of around 70 per cent were achieved for register.

Rhythm

Using different rhythms can create a large number of distinguishable motifs. Rhythm is a very effective parameter. Deutsch (1980) says that rhythm is one of the most important characteristics of a sound and Sumikawa, Blattner, Joy & Greenberg (1986) describe this as the most prominent characteristic of a motive. Sumikawa (1985) suggested that only seven time divisions should be used when creating rhythms. She also said that motives should be no longer than three or four notes or they will become too long for the user to easily remember and may also take too much time to play when in combination. Brewster, Wright & Edwards (1993) suggested that rhythms should be made as different as possible. Putting different numbers of notes in each rhythm is very effective. Patterson (1982) says that sounds are likely to be confused if the rhythms are similar even if there are large spectral differences. Small note lengths might not be noticed so notes less than sixteenth notes or semi-quavers should not be used. This depends on the tempo. If 180 beats per minute is used then sixteenth notes last 0.0825 seconds. If the sounds used are simple, for example just indicating events, then durations can be less. Short durations of only 0.03 seconds were shown to be usable and easily recognisable by listeners because they only communicated one thing. These short sounds help the earcons keep up with pace of interactions. McGookin (2004) found that users were able to achieve absolute identification rates of over 90 per cent for rhythm.

Dynamics

This is the change in volume of the motif. It can be made to increase as the motive plays (crescendo) or decrease (decrescendo). A crescendo could be used to give the idea of zooming a window, for example. Sumikawa (1985) suggest a total of five dynamics (soft, medium, loud, soft to loud and loud to soft). Brewster, Wright & Edwards (1993) says that care must be taken in the use of intensity. The overall sound level will be under the control of the user of the system. Earcons should all be kept within a close range so that if the user changes the volume of the system no sound will be lost. Some suggested ranges from Patterson (1982) are maximum 20 dB above threshold and minimum 10 dB above threshold.

One important ability of the human auditory system is that of auditory habituation (Buxton, Gaver & Bly, 1991) where continuous sounds with a restricted loudness range can fade into the background of consciousness after a short period of time. If the sound was to change or stop then it would come to the foreground of attention because of the sensitivity of the auditory system to changes in stimulus (Buxton, 1989).

Spatial Location

Another important factor in the perception of sound is spatial location. This is the ability to identify the position of a sound source in space. If a sound source is located to one side of the head, then the sound reaching the further ear will be reduced in intensity due to interaural intensity difference (IID) and delayed in time due to interaural time difference (ITD). This is known as the Duplex Theory and was first developed by Lord Rayleigh in 1907 as described by Moore (1989) and Levitt & Voroba (1974). IID and ITD can be used in auditory interfaces to provide directional or positional information.

Moore (1989) suggests that there can be an interaural intensity difference (IID) of up to as much as 20 dB across the ears. It can be seen that IID has little effect at low frequencies but is much more important for localisation at higher frequencies above 3 kHz. Interaural time difference (ITD), on the other hand, is only useful for localisation at lower frequencies because as frequencies increase to above 1500 Hz, the wavelength is shorter than the distance between the ears. When the sound source is at 90 degrees to the head (opposite one ear), there is a greater than 0.6 milliseconds delay between the ears. Delays of up to 2 milliseconds can be heard but longer than this the sound tends to be heard as two separate tones. Scharf & Houtsma (1986) provide more details on IID and ITD.

Humans can detect small changes in the position of a sound source. The minimum auditory angle (MAA) is the smallest separation between two sources that can be reliably detected. Strybel, Manligas & Perrott (1992) suggest that in the median plane sound sources only 1 degree apart can be detected. The median plane cuts through the head vertically, along the line of the nose. At 90 degrees azimuth (directly opposite one ear) sources must be 40 degrees apart. This has important implications for auditory displays. It indicates that high-resolution sounds can be used when presented in front of the user.

Much of the recorded sound we hear is in stereo. As Burgess (1992) describes, along with the sound, a stereo recording captures differences in intensity. The perceived position is along a line between the speakers. This simple, inexpensive technique can give useful spatial cues at the auditory interface. Pitt & Edwards (1991) have shown that using IID a user can find targets on an auditory display accurately and with reasonable speed. A listener wearing headphones uses lateralisation to locate the position of a sound which will be perceived as being on a plane between the ears and within the head (Burgess, 1992). Work has been done by Sakamoto, Gotoh & Kimaura (1976), Wenzel, Foster, Wightman & Kistler (1989), Gerhing & Morgan (1990) and Wenzel, Wightman & Kistler (1991) to make sounds perceived through headphones appear fully in three dimensions, or outside the head.

Spatial location has been used in earcon design to help differentiate multiple parallel earcons presented simultaneously (McGookin, 2002) and (McGookin & Brewster, 2004). It can also be used with serial earcons. Brewster, Wright & Edwards (1995) suggested that different families of earcons could be presented to different locations but this has not been investigated.

Combinations

One of the most powerful features of earcons is that they can be combined to produce complex audio messages. Blattner, Sumikawa & Greenberg (1989) say that the eloquence of motives lies in their ability to be combined to create larger recognisable structures. The repetition of motives, either exact or varied, or the linking of several different motives produces larger, more self sufficient patterns. These larger structures were used for earcons.

Sumikawa (1985) suggested some principles to keep in mind when constructing an earcon from motives: It should convey one basic meaning, be brief, simple and distinct from other earcons, and be easy to remember, identify and understand. She suggested three ways in which motives could be manipulated to create earcons. Repetition: Exact restatement of a preceding motive and its parameters. Variation: Altering one or more of the variable parameters from the preceding motive. Contrast: A decided difference in the pitch and / or rhythmic content from the preceding motive. These manipulations can be used in the following different ways.

Serial Compound Earcons

Serial compound earcons are the simplest type of complex audio messages and they are formed by placing two or more audio elements in succession. One motive is simply followed by another. This provides a simple and effective method for building up complex messages in sound. Brewster, Wright & Edwards (1993) suggested that, when playing earcons one after another, a gap between them should be used so that users can tell where one finishes and the other starts. A delay of 0.1 seconds is adequate. If the above guidelines are followed for each of the earcons to be combined then recognition rates should be similar to that of individual earcons.

The main drawback of serial compound earcons as proposed by Blattner, Sumikawa & Greenberg (1989) is that they can take a long time to play because each motive lasts a particular length of time depending on its notes and the tempo and these are then combined to produce longer compound earcons. If they were to be used in human computer interfaces, might not be able to keep up with the pace of interactions. Serial compound earcons could be played more rapidly (at a faster tempo) to overcome this problem but then errors in recognition may occur.

Parallel Compound Earcons

Parallel compound earcons are a method to reduce the length of time a compound audio message takes by playing the sequential component parts of a compound earcon in parallel so that they only take the time of one to play. Sound takes place sequentially in time. Blattner, Papp & Glinert (1992) say that our awareness and comprehension of the auditory world around us for the most part is done in parallel. This suggests that parallel compound earcons could exploit a natural ability of the human auditory system. Parallel compound earcons also use some of the attributes of the musical theory of counterpoint defined by Scholes (1975). In counterpoint individual instruments play separate musical lines which come together to make a musical whole. With parallel compound earcons, each component earcon is separate but the whole combined sound gives the meaning.

Another factor that may give parallel compound earcons an advantage over serial ones is the recency effect (Badeley, 1990), the term used to describe the enhanced recall of the most recently presented items. As serial earcons have to be held in memory for longer (because they take twice as long to play) the first part of the compound earcon might be forgotten and only the second, more recent earcon remembered. This problem would become more pronounced with longer earcons. Parallel compound earcons have all parts played at the same time so there is less time to forget any one earcon.

Brewster, Wright & Edwards (1993) conducted another experiment to see if the recognition and recall rates of parallel compound earcons was as accurate as the recall of serial compound earcons. His results show that parallel compound earcons are as easily recognised, effective and capable as serial compound earcons at communicating information. This indicated that parallel earcons are an effective means of reducing the length of serial compound earcons without compromising recognition rates. This means that parallel earcons are more effective in an auditory human computer interface as they take only the time of one to present to the user. This means that displaying complex information in sound that can keep pace with interactions is possible using parallel compound earcons. This allows a wider application of sound at the interface. His results also indicated that the more earcons were heard the better the recognition rates would be. This indicated that, if earcons were used at the human computer interface, then regular exposure would quickly lead to high levels of recognition. Musicians have again been shown to be no better than non-musicians. This means that auditory interfaces will be usable by most users whatever their level of musical skill.

Inherited Earcons

Inherited, family or hierarchical earcons provide a more powerful, hierarchical system. Each earcon is a node on a tree and inherits all the properties of the earcons above it. There is a maximum of six levels to the tree as there are only six parameters that can be varied (rhythm, pitch, timbre, register, dynamics and spatial location). With the inheritance hierarchy there is only one new piece of information to remember at each level, thus making family earcons easier to remember. Initially, there will be much to learn but later extensions will be simple.

Transformed Earcons

Transformational earcons are another type of family earcons. In transformational earcons each data attribute is mapped to a parameter of the earcon. For example, if transformational earcons were used to represent files in a computer interface, the file type could be represented by rhythm, size by the timbre used, and creation date by the register. Each file type would be mapped to a unique rhythm. Therefore, two files of the same type and same size, but different creation date, would share the same rhythm and be played by the same timbre, but would be presented in a different register. If two files were of different types, but were of the same size and shared the same creation date, they would be represented by different rhythms but would be played by the same timbre and in the same register.

Memorability

An important factor in the usability of earcons is memorability. If they become too long this may be a problem so Sumikawa (1985) suggests that they should be kept as short as is necessary to get their message across. Deutsch (1980) suggests that structured sequences of tones should be more easily remembered than random tones. The strong rhythmic structure will also aid memory. Earcons have some advantages over auditory icons because they have a strong structure to link them together. This may reduce the learning time and the memory load.

Conclusions

Earcons have been shown to be a very effective method of communicating complex information in non-verbal sounds and they can even do it at a rate which can keep up with the pace of interactions in a user interface. Since the mapping between the earcon and what it represents is abstract, earcons can cover more content domains more flexibly than auditory icons and can represent any possible concepts. In a user interface for non-visual presentation, navigation and manipulation of structured documents, for example, different element types like sections, headings, paragraphs, lists or tables or different operations like focus, select, activate, move, remove and insert.

2.2.1.3 Spearcons

Spearcons or speech-based earcons were developed by Walker, Nance & Lindsay (2006). They describe spearcons as brief audio cues that can play the same roles as earcons and auditory icons, but in a more effective manner, overall.

Spearcons are created automatically by converting the text description of an item, for example “Heading”, to speech via text-to-speech (TTS), and then speeding up the resulting audio clip (without changing pitch) to the point that it is no longer comprehensible as speech. Spearcons are unique to the specific item, just as with auditory icons, although the uniqueness is acoustic, and not semantic or metaphorical. These unique sounds are analogous to fingerprints, a unique identifier that is only part of the information contained in the original, because of the acoustic relation between the spearcons and the original speech phrases. At the same time, the similarities in the item content cause the spearcons to form families of sounds. For example, the spearcons for “Table”, “Table Row”, and “Table Cell” are all unique, including being of different lengths. However, they are acoustically similar at the beginning of the sounds, which allows them to be grouped together even though they are not comprehensible as any particular words. The different lengths help the listener learn the mappings, and provide a guide to the ear while scanning through different items.

Since the mapping between spearcons and their items is non-arbitrary, there is less learning required than would be the case for a purely arbitrary mapping as with earcons described in the previous chapter. Spearcons provide more direct mappings between sounds and items than earcons and cover more content domains, more flexibly, than auditory icons. Spearcons can be created algorithmically, although some hand tweaking is sometimes preferable, so they can be created dynamically as needed, even on the fly, and can represent any possible concept. This allows the system to cope with dynamically changing items. Also, spearcons are easy to learn whether they are comprehensible as a particular word or not, because they derive from the original speech (Palladino & Walker, 2007).

Spearcons have recently shown promising results in menu navigation tasks. Walker, Nance & Lindsay (2006) demonstrated that adding spearcons to a text-to-speech menu leads to faster and more accurate navigation than text-to-speech only, auditory icons + text-to-speech, and earcons + text-to-speech groups. Spearcons also improved navigational efficiency more than menus using only text-to-speech or no sound when combined with visual cues (Palladino & Walker, 2008). According to Palladino & Walker (2008), in their visuals-off condition, the mean time-to-target with spearcons + text-to-speech is shorter than that with text-to-speech only, despite the fact that adding spearcons makes the total system feedback longer.

2.2.1.4 Synthetic Speech

Presenting information in speech is slow because of its serial nature. To assimilate information the user must hear it from beginning to end and many words may have to be comprehended before a message can be understood. Speech suffers from many of the same problems as text in text-based computer systems, as this is also a serial medium.

Work has been done on increasing the presentation rate of synthetic speech by Aldrich & Parkin (1989) and Slowiaczek & Nusbaum (1985). Both of these found that the accuracy of recognition decreased as the speech rate went up. Slowiaczek & Nusbaum (1985) found that a speaking rate of about 150 words per minute is optimal for perception of synthetic speech. This figure is around the normal speaking rate and is very slow to listen to. When they increased the rate to 250 words per minute (the normal sight-reading rate) recognition decreased significantly. One of the main causes of this, they suggest, is the poor quality of the synthetic speech. Much of the prosodic information (intonation, pausing, etc.) in normal speech is not given in synthetic speech. At low speeds listeners can cope without it but at higher speeds it becomes much more important. With these problems, users of synthetic speech will be constrained to operate at rates that are far below normal reading rates. Highly-skilled users, such as blind and visually impaired persons, can reach higher recognition rates but this requires much practice.

Synthetic speech is mainly purposed for presenting text. However by manipulating attributes of the speech like the voice, pitch, register, tempo and spatial location it is also possible to encode some non-text data attributes along with a text.

Sorin, Lemarié, Aussenac-Gilles, Mojahid & Oriola (2014) used specialized audio and two different voices to demarcate headings in their work on communicating text structure to blind persons with text-to-speech. Their research showed that text comprehension was slightly improved but failed to show statistically significant evidences that either the dual-voices method or the spatialised audio method has better performance than current text-to-speech text description saying “Heading level N” before the heading oralisation. However, their document structures were very simple, only consisting of two different element types, namely headings and non-heading document contents. If more complex document structures were used then a performance increase may have been found.

Smither (1993) conducted an experiment to investigate the demands synthetic speech puts on short term memory. He tested natural speech against synthetic speech on young and old adults. His results showed that synthetic speech put a heavier load on short term memory than natural speech. Older subjects performed worse than younger ones but both groups performed worse with the synthetic speech.

One important ability of the human auditory system is that of auditory habituation (Buxton, Gaver & Bly, 1991) where continuous sounds with a restricted loudness range can fade into the background of consciousness after a short period of time. If the sound was to change or stop then it would come to the foreground of attention because of the sensitivity of the auditory system to changes in stimulus (Buxton, 1989). Auditory habituation is difficult to achieve with speech because of the large dynamic range it uses. According to Patterson (1982) the vowels of speech are often 30 dB more intense than the consonants, and so, if a voice message were attenuated to produce a background version with the correct vowel level the consonants would be near or below masked threshold. Also a voice message could not continuously speak its text without problems of masking.

The use of speech may also be annoying for other users who overhear the interface. Hapeshi & Jones (1992) say that a number of studies have demonstrated that these attention grabbing qualities of the auditory channel, combined with its special adaptation to speech, makes it highly disruptive when listeners are performing some verbal activity. They also give evidence to show that background speech, even at low intensities, is much more disruptive than non-speech sound when recalling information.

Badeley (1990) reports the unattended speech effect. Unattended speech in the background causes information to be knocked out of short-term memory, whereas noise or non-speech sound does not even when it is pulsed to give the same intensity envelope as speech. This problem is unaffected by the intensity of the speech, provided that it is audible. This shows a problem for speech at the interface as it is likely to prove disruptive for other users in the same environment unless it is kept at a very low intensity which can cause problems with the ability to hear consonants.

2.2.2 Haptic Approaches

In the haptic domain one common form of text-based output is Braille in addition to Vibratese and the iconic counterparts are tactons and hapticons.

2.2.2.1 Tactons

Tactons, or tactile icons, had been developed by Brewster & Brown (2004). These are structured, abstract vibrotactile messages that can be used for presenting multidimensional information non-visually.

Tactons are created by encoding information using the parameters of cutaneous perception. Cutaneous perception refers to the mechanoreceptors contained within the skin, and includes the sensations of vibration, temperature, pain and indentation. Tactile devices are used to present feedback to the cutaneous sense. As tactons are abstract the mapping between the tacton and what it represents must be learned, but work on earcons has shown that learning can take place quickly (Brewster, 1998).

The encoding is similar to that of earcons in sound as proposed by Blattner, Sumikawa & Greenberg (1989) where each of the musical parameters, for example timbre, pitch or register is varied to encode information. Similar parameters can be used for tactons although their relative importance is different. The properties that can be manipulated for tactons are similar to those used in the creation of earcons. The parameters for manipulation also vary depending on the type of transducer used. Not all transducers allow all types of parameters to be manipulated.

According to Brown (2007), mobile and wearable devices commonly have a very simple single point-contact tactile stimulator built-in that can alert the user to a call. These are small DC motors featuring an eccentric weight in the shaft. When they are switched on the shaft spins, and the spinning of the eccentric weight causes the sensation of vibration. These vibration motors typically vibrate at frequencies around 130 Hz. With a phone motor the whole casing in which the motor is housed vibrates. Therefore, when this motor is mounted in a mobile or wearable device, the whole body of the device vibrates. In addition, less control over the stimuli is available. Only the on-off times and the velocity of the rotation can be controlled. This enables them to produce pulses of different durations and intensities. Therefore, when using phone vibration motors, tactons must be designed within these constraints.

Frequency

A range of frequencies can be used to differentiate tactons. The range of 20 - 1000 Hz is perceivable but maximum sensitivity occurs around 250 Hz (Gunther, Davenport & O'Modhrain, 2002). The number of discrete values that can be differentiated is not well understood, but Gill (2003) suggests that a maximum of nine different levels can be used. As in audition, a change in amplitude leads to a change in the perception of frequency so this has an impact on the use of frequency as a cue. The number of levels of frequency that can be discriminated also depends on whether the cues are presented in a relative or absolute way. Making relative comparisons between stimuli is much easier than absolute identification, which will lead to much fewer discriminable values, as shown in the work on earcon design (Brewster, Wright & Edwards, 1992).

Amplitude

The amplitude of stimulation can also be used to encode values to present information to the user. Gunther, Davenport & O'Modhrain (2002) reports that the intensity range of the skin extends to 55 dB above the threshold of detection, beyond which vibrations may become unpleasant or painful (Vitense, Jacko & Emery, 2003). Craig & Sherrick (1982) indicate that perception deteriorates above 28 dB so this would seem to be a useful maximum. Gunther (2001) reports that various values, ranging from 0.4 dB to 3.2 dB, have been reported for the just noticeable difference (JND) value for intensity. Gill (2003) states that that no more than four different intensities should be used. Again the number of useful discriminable values will depend on absolute or relative presentation of stimuli.

Intensity

Due to the interactions between frequency and amplitude several researchers have suggested that they be combined into a single parameter to simplify design. Sherrick & Cholewiak (1986) found that combining frequency and amplitude redundantly allowed a greater number of identifiable levels to be created. Brown (2007) says that, when using a phone vibration motor, the intensity of the vibration is changed by adjusting the supply voltage to the phone motor. When the supply voltage of a phone motor is adjusted it changes both the amplitude and the frequency of the resulting vibration. This, therefore, provides redundant coding of the same information in both of these parameters of vibration and may help people to identify multiple levels.

Waveform

The perception of wave shape is much more limited than with the perception of timbre in sound. Users can differentiate sine waves and square waves but more subtle differences are more difficult (Gunther, 2001). This limits the number of different values that can be encoded and makes this a much less important variable than it is in earcon design where it is one of the key variables. However, an experiment conducted by Hoggan & Brewster (2007) showed that users can identify and distinguish differing waveforms significantly more effectively than different frequencies and tactile roughness (amplitude modulation). Therefore, different waveforms should be used as the texture parameter in tacton design.

Roughness

Due to the problems of recognition of simple waveforms such as square waves and sine waves it is necessary to consider more complex waveforms. Experimentation by Brown & Brewster (2005) and (2006) indicated that perceptually different waveforms could be created using amplitude modulated sinusoids. These are created by multiplying a sine wave of a given frequency by a sine wave of another frequency, thus modulating the amplitude of the base signal by the second frequency.

Dynamics

Tactile versions of musical dynamics can be created by manipulating the amplitude of vibrations to create increasing (crescendo), decreasing (decrescendo), and level stimuli. Brown & Brewster (2006) carried out an experiment to test perception of these stimuli. Identification rates of 92 to 100 per cent indicate that these tactile dynamics can be identified and distinguished from each other, and that tactile dynamics could be used in tacton design.

Duration

Pulses of different durations can be used to encode information. Gunther (2001) investigated a range of subjective responses to pulses of different durations. He found that stimuli lasting less than 0.1 seconds were perceived as taps or jabs whereas stimuli of longer duration, when combined with gradual attacks and decays, may be perceived as smoothly flowing tactile phrases. He suggests combining duration with alterations in the envelope of a vibration, for example an abrupt attack feels like a tap against the skin, a gradual attack feels like something rising up out of the skin. Building on from duration, groups of pulses of different durations can be composed into rhythmic units and the tempo at which these rhythms are played can also be varied.

Rhythm

Using different rhythms can create a large number of distinguishable motifs. Rhythm is a very effective parameter in both sound and touch. Deutsch (1980) says that rhythm is one of the most important characteristics of a sound and Sumikawa, Blattner, Joy & Greenberg (1986) describe this as the most prominent characteristic of a motive. Sumikawa (1985) suggested that only seven time divisions should be used when creating rhythms. She also said that motives should be no longer than three or four notes or they will become too long for the user to easily remember and may also take too much time to play when in combination. Brewster, Wright & Edwards (1993) suggested that rhythms should be made as different as possible. Putting different numbers of notes in each rhythm is very effective. Patterson (1982) says that sounds are likely to be confused if the rhythms are similar even if there are large spectral differences. Small note lengths might not be noticed so notes less than sixteenth notes or semi-quavers should not be used. This depends on the tempo. If 180 beats per minute is used then sixteenth notes last 0.0825 seconds. If the sounds used are simple, for example just indicating events, then durations can be less. Short durations of only 0.03 seconds were shown to be usable and easily recognisable by listeners because they only communicated one thing. These short sounds help the earcons keep up with pace of interactions. McGookin (2004) found that users were able to achieve absolute identification rates of over 90 per cent for rhythm.

Tempo

Brewster, Wright & Edwards (1995) said that changing the tempo, speeding up or slowing down the sounds, is another effective method for differentiating earcons. In addition to Brewster's guidelines also advice given by Van Erp & Spapé (2003) identified tempo (speed) as an important parameter in the identification of tactile melodies. The tempo is expressed in beats per minute (bpm).

Spatial Location

Spatially distributed transducers can encode information in the position of stimulation across the body. The choice of body location for vibrotactile display is important, as different locations have different levels of sensitivity and spatial acuity. Certain body locations are particularly suitable, or particularly unsuitable, for certain types of vibrotactile displays. The fingers are often used for vibrotactile displays because of their high sensitivity to small amplitudes and their high spatial acuity (Craig & Sherrick, 1982). However, the fingers are often required for other tasks, so other body locations may be more suitable. Craig & Sherrick (1982) suggest the back, thigh and abdomen as other suitable body locations. Transducers should not be placed on or near the head, as this can cause leakage of vibrations into the ears, resulting in unwanted sounds (Gunther, Davenport & O'Modhrain, 2002).

Combinations

As suggested by Blattner, Sumikawa & Greenberg (1989), short motifs could be used to represent simple objects or operations and these can then be combined in different ways to represent more complex tactile messages and concepts.

Serial Compound Tactons

Serial compound tactons are the simplest type of complex tactile messages and they are formed by placing two or more tactile elements in succession. One motive is simply followed by another. This provides a simple and effective method for building up complex messages in vibration. Brewster, Wright & Edwards (1993) suggested that, when playing motives one after another, a gap between them should be used so that users can tell where one finishes and the other starts. A delay of 0.1 seconds is adequate. If the above guidelines are followed for each of the tactons to be combined then recognition rates should be similar to that of individual tactons.

Inherited Tactons

As with inherited, family or hierarchical earcons, tactons can also be combined in a hierarchical way. Each tacton is a node in a tree and inherits properties from the levels above it. With the inheritance hierarchy there is only one new piece of information to remember at each level, thus making family tactons easier to remember. Initially, there will be much to learn but later extensions will be simple.

Transformed Tactons

A third type of combined tactons is the transformational tacton. These have several properties, each represented by a different tactile parameter. For example, if transformational tactons were used to represent files in a computer interface, the file type could be represented by rhythm, size by frequency, and creation date by body location. Each file type would be mapped to a unique rhythm. Therefore, two files of the same type, and same size, but different creation date would share the same rhythm and frequency, but would be presented to a different body location. If two files were of different types but the same size they would be represented by different rhythms with the same frequency.

2.2.2.2 Hapticons

MacLean & Enriquez (2003) recently proposed Hapticons or haptic icons, which they define as brief programmed forces applied to a user through a haptic interface, with the role of communicating a simple idea in a manner similar to visual or auditory icons. These use kinesthetic one degree-of-freedom force-feedback devices, rather than tactile displays, so encode information very differently to tactons described in the previous Chapter 2.2.5.

MacLean & Enriquez (2003) used multidimensional scaling techniques (MDS) to determine how haptic icons can be created from signal parameters such as waveform, frequency, and force. They found that for the ranges of parameters that they implemented in a handheld knob, frequency played a dominant role in distinguishing between the multidimensional stimuli and that waveform and force were less salient.

Hapticons are formed from basic building blocks of haptic communication called haptic phonemes. A haptic phoneme represents the smallest unit of a constructed haptic signal to which a meaning can be assigned with variations in waveform and frequency. These haptic phonemes can be combined serially or in parallel to form haptic words, or haptic icons, which can hold more elaborate meanings for their users. With this method, Enriquez, MacLean & Chita (2006) created a set of 9 haptic icons that varied in terms of waveform and frequency. They then trained participants to associate each haptic icon with an arbitrary concept, such as the name of a fruit. They found that participants learned these associations after about 25 minutes of training and achieved higher identification rates of 81 per cent correct with stimuli that varied in frequency, compared to those that varied in waveform with 73 per cent correct.

Enriquez & MacLean (2003) designed and implemented the Hapticon Editor, which is a tool and graphical interface for the creation and editing of Hapticons. The tool's features include various methods for creating new icons including direct recording of manual trajectories and creation from a choice of basis waveforms, novel direct-manipulation icon editing mechanisms, integrated playback and convenient storage of icons to a file.

There are no kinesthetic one degree-of-freedom force-feedback devices embedded in mobile and wearable devices because of their large size and weight. Therefore hapticons cannot be used in this research.

2.2.2.3 Braille

Braille is a very common method of tactile output used by blind and visually impaired persons (Foulke, 1982). Many different dynamic Braille displays are available. A Braille display is made up of a line of soft cells (often 40 or 80), each with 6 or 8 pins that move up and down to represent the dots of a Braille cell. The user can read a line of Braille cells by touching the pins of each cell as they pop up. The cells are very low resolution using a maximum of 8 pins. These displays are also very expensive.

The mappings between characters and the tactile characters must be learned. These mappings are mainly based on a numerical model where each Braille character consists of a three row by two column cell. These combinations of raised dots allowing 64 individual patterns to be displayed. The patterns represent the letters of the alphabet and punctuation. Many empirical evaluations have been conducted that indicate that Braille code is more effective than using embossed letters in terms of reading speed and text recognition (Sharmin, Evreinov & Raisamo, 2005). The main disadvantage of Braille is that very few blind and visually impaired persons can read it. For example, it has been estimated that only 2 per cent of British blind persons are able to reading Braille (Bruce & McKennell, 1991). Braille tends to be used mainly for representing text although other notations are used, for example for music.

Minatani (2012) proposed a way of realising advantages attainable from making use of the logical structure of documents by developing a method of presenting the tree structure information of a document on a Braille display in his research on document structure presentation and navigation using a Braille display. The document browser software developed for this research operates as a DAISY player. Experimentation found that using a user interface of that document browser software improves the efficiency of understanding the document's general structure and finding headings when compared to the user interface of a conventional DAISY player with numeric keypad cursor navigation.

There are no dynamic Braille displays embedded in mobile and wearable devices because of their large size and weight. Therefore Braille cannot be used in this research.

2.2.2.4 Vibratese

Geldard (1957) developed a vibrotactile language, called Vibratese, which varied three parameters (duration, intensity and spatial location) in order to assign a unique vibrotactile stimulus to every letter of the alphabet. With Vibratese, text is encoded very differently to Braille, described in the previous Chapter, where the user actively has to touch the pins of the Braille cells as they pop up.

Three durations and three intensities were used, and five vibrotactile transducers were positioned in an X-pattern on the user's chest with only one transducer activated at a time. These parameters were manipulated independently, resulting in 45 unique combinations of these three parameters. The most commonly occurring letters were represented by the shortest durations to speed up presentation. The encoding scheme used in Vibratese is another example of the coded approach, as the mapping is abstract. As with Braille, there is no direct relationship between the vibrotactile stimulus and the letter it represents.

This language was found to be quite successful. After several training sessions, one subject managed to interpret this language at a rate of 35 words per minute. The results achieved with the Vibratese language indicate that vibrotactile display holds promise for communicating complex information, and that using an abstract approach to encoding information in vibrotactile messages can be successful.

2.2.3 Crossmodal Icons

Hoggan & Brewster (2006) introduced the concept of crossmodal icons. These are abstract icons which can be automatically instantiated in one of two equivalent forms (auditory or tactile) as either an earcon or a tacton, such that the resultant earcons or tactons are intuitively equivalent and can be compared as such. Crossmodal icons enable the same information to be presented interchangeably via different modalities.

Crossmodal interaction is a subset of multimodal interaction where the different senses are used to receive the same data. This provides a common representation of the data from both senses, in this case audio and tactile, (Gibson, 1966) making them congruent informationally (Marks, 1978). Crossmodal use of the different senses allows the characteristics of one sensory modality to be transformed into stimuli for another sensory modality. Multimodal interaction, on the other hand, may also use the different senses to receive different information.

Auditory and tactile displays were chosen because they are ideal candidates for crossmodal combination in view of the fact that our senses of hearing and touch share several important similarities, in particular their temporal and spatial characteristics and their ability to perceive vibrations. It has been suggested that the more properties shared between two modalities, the stronger will be the observer's unity assumption that information from different sensory channels can be attributed to the same distal event or object (Adelstein, Begault, Anderson & Wenzel, 2003).

One important area of study is the set of parameters that can be used to create crossmodal icons where the same information can be easily mapped between the two modalities. If there is no direct mapping between modalities, an abstract mapping must be developed where the cues may still be perceived as equivalent representations of information. An attribute that can communicate comparable information across modalities is considered to be amodal (Lewkowicz, 2000). The following amodal parameters have been derived from a survey of related work on the parameters available in the earcon (Brewster, Wright & Edwards, 1993) and tacton (Brown, Brewster & Purchase, 2005) domain, which have been derived from psychoacoustics and psychophysics.

Rhythm

As outlined by Brewster, Wright & Edwards (1993), short rhythms can be used to represent objects or actions. Such rhythms can be both audio and tactile due to their shared temporal properties. The earcon would play a rhythm by a series of notes of different durations via a loudspeaker. The corresponding tacton would transmit this same rhythm via a series of pulses through a vibrotactile display.

Roughness

Roughness can be mapped between modalities using amplitude modulation in the vibrotactile cues and differing timbres in the audio domain. Modulating the amplitude of a tactile pulse creates differing levels of roughness. A preliminary experiment was conducted in order to determine which version of audio roughness can be perceived as equivalent and maps most effectively to tactile roughness. Initial results show that participants preferred the use of differing timbres (for example flute or tremolo strings) in audio to represent the different levels of roughness used in tactile. However, the results also show that there is no significant difference in performance between timbre and audio amplitude modulation. Therefore, amplitude modulation or timbre in audio can be perceived as equivalent and map effectively to tactile roughness but timbre is the preferred choice.

Intensity

Intensity is another directly transferable parameter between earcons and tactons. High or low intensity could be achieved by increasing or decreasing the volume of the earcon or the intensity of the corresponding tacton. However, earcon guidelines suggest that it should be used vary carefully as it can cause annoyance and has few absolutely discriminable values (Brewster, Wright & Edwards, 1993). Further studies are required to determine the usefulness of this parameter.

Spatial Location

Unlike rhythm and intensity, spatial location cannot be directly transferred from the tactile domain to the audio domain because the spatial location of transducers placed on the body is concrete while spatial location in audio environments is an abstract concept. Research is needed to investigate how to map from a tactile location on the body to an audio location in a soundspace. For example, one possibility is to place the tactile transducers around the waist whilst using an audio display presented through headphones. Then, a location on the left could be presented via an earcon by panning the audio to the left of the soundspace and tactons could give the same cue by activating a transducer placed on the left hand side of the waist.

Evaluation

Hoggan & Brewster (2007) conducted an experiment to investigate absolute identification of audio and tactile crossmodal icons when a user is trained in one modality and tested in the other (and given no training in the other modality) to see if knowledge could be transferred between modalities. The performance when users were static and mobile had also been compared to see any effects that mobility might have on the recognition of the cues. A complete set of crossmodal icons was created by encoding three dimensions of information in three crossmodal auditory and tactile parameters.

The results showed that if participants were trained in sound with earcons and then tested with the same messages presented via tactons they could recognize 85 per cent of messages when stationary and 79 per cent when mobile. When trained with tactons and tested with earcons participants could accurately recognize 76.5 per cent of messages when stationary and 78 per cent of messages when mobile. These results suggest that participants can recognize and understand a message in a different modality very effectively.

2.3 Assistive Technologies

The assistive technology dynamically accesses and updates the structure, content and style of the structured document provided by the accessibility API of the operating system and provides an alternative user interface for presentation, navigation and manipulation to the user by employing various different modalities. In addition to screen readers, there are many other research projects aiming to provide a solution for non-visual presentation, navigation and manipulation of different types of structured data.

Table 2.2 shows the different assistive technology approaches available and the modalities employed for presenting data serving the auditory and haptic human sense, as well as how they compare to each other, their differences and whether they are all of equal significance or whether some have made more impact, or seen greater usage, than others:

Modalities	Assistive Technologies
• **Synthetic Speech** • **Braille**	• **Screen Readers**
• **Auditory Icons**	• **SonicFinder** (Gaver, 1989) • **SharedARK** (Gaver & Smith, 1990) • **ARKola** (Gaver, Smith & O'Shea, 1991) • **RAVE** (Gaver, Moran, MacLean, Lövstrand, Dourish, Carter & Buxton, 1992) • **ShareMon** (Cohen, 1992) • **AB-Web** (Roth, Petrucci, Assimacopoulos & Pun (1998) • **WebSound** (Petrucci, Harth, Roth, Assimacopoulos & Pun, 2000)
• **Earcons**	• **TableVis** (Kildal & Brewster, 2006) • **Hearcons** (Donker, Klante & Gorny, 2002) • **Auditory Maps** (Blattner, Papp & Glinert, 1992) • **Algebra Earcons** (Stevens, Brewster, Wright & Edwards, 1994)
• **Tactons**	• **TactoWeb** (Petit, Dufresne & Robert, 2011) • **MaskGen** (Petit, Dufresne, Levesque, Hayward & Trudeau, 2008) • **Haptic Emoticons** (Rovers & Van Essen, 2004) • **Vibrocons** (Van Erp & Van Veen, 2001)
• **Crossmodal Icons**	• **CrossTrainer** (Hoggan & Brewster, 2010)

Table 2.2: Assistive Technologies and different Modalities employed

2.3.1 Screen Readers

A screen reader gets information about what is displayed on the screen from the accessibility API of the operating system and presents this content to the user by employing different non-visual modalities like synthetic speech or Braille. Screen Readers are useful for persons who are blind, suffer from a severe visual impairment as well as illiterate or persons with a learning disability. Many different screen reader products are available on the market, which can either be proprietary or built into the operating system. They can be categorised into commercial as well as free or open source solutions.

The screen readers available for desktop computers are a mix of different products: Freedom Scientific [JAWS] is a commercial screen reader for Microsoft Windows which can be used in combination with the [MAGic] screen magnifier from the same company. This is one of the most common used product because of the large market share of Microsoft Windows as an operating system for desktop computers. Aisquared [ZoomText Reader] is a basic commercial screen reader for the Microsoft Windows operating system too, optimised for interworking with the [ZoomText] screen magnifier from the same manufacturer. Microsoft [Windows Reader] is a very basic screen reader that comes with the Microsoft Windows operating system free of charge. Apple [VoiceOver] is the screen reader included free of charge in Apple's MacOS operating system. Dolphin [Supernova] is a commercial screen reader for the Microsoft Windows operating system, designed for use in combination with the [Lunar] screen magnifier manufactured by the same company. Nonvisual Desktop Access [NVDA] is a free and open source screen reader project for the Microsoft Windows platform. GWMicro [Windows Eyes] is a commercial screen reader for the Microsoft Windows operating system very frequently used in the United States.

The screen readers available on mobile and wearable devices are a mix of different products too: Apple [VoiceOver for iOS] is the screen reader included in Apple's iOS operating system free of charge. This is one of the products most used because Apple's iDevices are very popular with blind and visually impaired persons due to the fact that Apple was the first manufacturer, providing assistive technologies like screen reader and screen magnifier out of the box and free of charge. Google [TalkBack] is the screen reader that comes with the Google [Android] operating system free of charge. Nuance [Talks] is a commercial screen reader for Nokia's Symbian operating system. CodeFactory [MobileSpeak] is a commercial screen reader available for both, the Nokia SymbianOS as well as for the Microsoft Windows Mobile operating system.

2.3.2 SonicFinder

The SonicFinder was developed by Gaver (1989) from the ideas of auditory icons, which was one of the first attempts to actually integrate non-speech audio into a human-computer interface. This is an interface that runs on the Apple Macintosh alongside the ordinary Finder and provides auditory representations of some objects and actions within the interface. This system is not designed for blind users but as an aid for Mac users without disability. Files are given a wooden sound, applications a metal sound and folders a papery sound. The larger the object the deeper the sound it makes. Thus, selecting an application means tapping it. It will make a metal sound which will confirm that it is an application and the deepness of the sound will indicate its size. Copying uses the idea of pouring liquid into a receptacle. The rising of the pitch indicates that the receptacle is getting fuller and when it is full the copy is complete. To delete a file on the Macintosh it is dragged and dropped into the waste basket. This makes the sound of smashing dishes to indicate destruction.

2.3.3 SharedARK and ARKola

Gaver & Smith (1990) extended his ideas of auditory icons into the area of large-scale collaborative environments. They based their work around the Shared Alternative Reality Kit (SharedARK). In this system a virtual physics laboratory was modelled. Multiple users could perform virtual experiments on objects in an environment that extends far beyond the view offered by their screens. Three groups of sounds were used: Confirmatory sounds, process and state sounds and navigation aids.

The confirmatory group of sounds used many of the principles from the SonicFinder. Clicking on a button made a tapping sound or putting one object on top of another made a wooden sound. The process and state information sounds have two levels. There are global states for the whole system and the sounds representing these can be heard everywhere within the system. These sounds are designed to fade into the background of consciousness. Other more localised sounds are then used for each of the specific experiments. If the user moves from one experiment to another, the new experiment's sound will be louder than that of the other, which is further away. This type of constant audio feedback is an important step forward from previous systems that just presented event information in sound. Much of the feedback displayed in graphical interfaces is constant, changing very little over time. Constant audio feedback is a useful extension to simple event feedback.

One of the most interesting concepts in SharedARK is the soundholder. These are auditory landmarks that users can navigate by. They can be placed anywhere in the system and constantly emit sounds whose volume decreases as the user moves away from them or increases as the user gets closer. Gaver & Smith (1990) suggest using environmental sounds such as bird calls or burbling streams for these, as they are very distinct and easy to remember.

In the ARKola system Gaver, Smith & O'Shea (1991) again used the SharedARK but this time a soft drinks factory was modelled. The simulation consisted of a set of nine machines. The machines were split into two groups: Those for input and those for output. The input machines supplied the raw materials, the output machines capped the bottles and sent them for shipping and financing. Each machine had an on / off switch and a rate control. The aim of the simulation was to run the plant as efficiently as possible, avoid waste of raw materials and make a profit by shipping bottles. Two subjects used the system at the same time from physically separate locations. They communicated via a two-way audio / video link.

Each of the machines had a sound to indicate its status over time, for example the bottle dispenser made the sound of clinking bottles. The rhythm of the sounds reflected the rate at which the machine was running. If a machine ran out of supplies or broke down its sound stopped. Sounds were also added to indicate that materials were being wasted. A splashing sound indicated that liquid was being spilled, the sound of smashing bottles indicated that bottles were being lost. The system was designed so that up to 14 different sounds could be played at once. To reduce the chance that all sounds would be playing simultaneously, sounds were pulsed once a second rather than playing continuously.

2.3.4 RAVE and ShareMon

The RAVE system, developed by Gaver, Moran, MacLean, Lövstrand, Dourish, Carter & Buxton (1992) at EuroPARC, was designed to allow physically separated colleagues to work together effectively and naturally. One example of this was the installation of two-way video cameras into offices so that people could communicate whilst still being physically separated. To deal with some of the issues of privacy this brought up, auditory icons were added. For example, when someone started to look into a particular office, a door opening sound was played to the people in that office so that they knew they were being looked at. When the connection was broken a door closing sound was played. Other sounds included: A knock or telephone bell to indicate a videophone request and footsteps to indicate sweeps, where one person briefly sweeps their video eyes over all the other offices to see who is around, similar to walking by and looking through the doors. These sounds provide a powerful way of communicating information to users without distracting their attention.

The ShareMon, proposed by Cohen (1992) and (1993), uses auditory icons to give information about background file sharing activity on the Apple Macintosh. Cohen wanted to present this information without disrupting foreground activities. ShareMon indicates when another user logs-on to your machine in a similar way to the RAVE system described above: A knocking sound is given when a user logs-on and a door slamming sound when they log-off. Other sounds include: A draw opening or closing sound to indicate open or close file, a chair creak to remind the user that there is still a connection in place and walking sounds to indicate level of copying activity, where walking means low activity and running means high activity.

2.3.5 TableVis

Kildal & Brewster (2006) designed an interface to obtain overview information from complex numerical data tables non-visually. The table sonification technique of Kildal & Brewster (2005), that hides detail in the data and highlights its main features without doing any computations to the data, is combined with a graphics tablet for focus and context interactive navigation, in an interface called TableVis.

Like in the work by Ramloll, Brewster, Yu & Riedel (2001), the value of a cell in the table is mapped to a pitch, with frequencies in the range of 27.5 Hz to 4.786 kHz (21 - 108 MIDI values). Higher values are represented by higher pitches and vice versa, which is accepted as natural and intuitive (Roffler & Butler, 1967). The location of the cell in the table is mapped to the spatial location of the earcon in stereo from left to right equidistant from each other, following the metaphor of mapping the values in the audio space around the user according to how they are located in the table. These sounds are panned so that all the values of the leftmost column are played on the left channel, all the values of the rightmost column on the right channel and the rest of the columns at regular distances between left and right. These earcons are generated using MIDI, and the timbre chosen is the piano, because of its rapid attack and wide pitch range.

A graphics tablet was chosen to control TableVis, as it offers a number of advantages. It is an absolute positioning pointing device. The borders of the working area of the tablet can be felt with both hands. The table to be explored is presented on the working surface of the tablet, scaled to fill it completely. Users can access any region of the table directly, and the distance of the pen relative to the borders informs them of which area of the table is being explored at any time, providing contextual information.

TableVis offers three modes of exploration. In the cells mode, the cell being pointed at on the tablet is sonified, facilitating freehand exploration of the data table. In rows and columns modes, a complete row or column is sonified when the pen is pointed at it. These two last modes facilitate obtaining quick overviews of the complete data set. It is enough to draw a line across the whole table, horizontally in columns mode or vertically in rows mode, to generate the data sonification interactively.

The values in a row or column are not sonified simultaneously, which would produce a generally dissonant chord of high musical complexity. Instead, they are sonified in very rapid succession, from left to right or from top to bottom. This approach takes advantage of the high temporal resolution of sound, a difference of 20 ms being enough to be able to tell which of two sounds is heard first (Pierce, 2001). For a good compromise between browsing speed and clarity, the duration of the sound corresponding to one cell was chosen to be 60 ms, avoiding overlapping between sounds of neighbouring cells.

Results from the evaluation of this interface suggest that this technique can deliver better scores than speech in time to answer overview questions, correctness of the answers and subjective workload.

2.3.6 AB-Web and WebSound

Roth, Petrucci, Assimacopoulos & Pun (1998) and Roth, Petrucci, Pun & Assimacopoulos (1999) developed a multi-modal auditory interface called AB-Web which transforms the two dimensional visual representation of a HTML document into a 3D immersive virtual audio navigable environment and permits blind users to work more easily and efficiently with GUI web-browsers. As with TableVis, a touch-sensitive screen is used to facilitate user interaction. Their hypothesis is that a 3D immersive virtual sound navigable environment combined with haptic manipulation of the audio environment can enable blind user to construct a mental representation of the spatial document.

Two successive exploration phases are performed by users in order to analyse a HTML document. First, a macro-analysis phase allows the understanding of the document structure and the element types (for example text, images and forms) displayed in the current browser viewport. Secondly, a micro-analysis phase lets users focus on one particular object and thus obtain its information content. These phases are iterated: Users may perform a succession of macro- and micro-analysis explorations of the document in order to understand it.

For marco-analysis, two complementary modes of interaction, passive and active, are provided. In the passive macro-analysis, the boundary location (in cm) and type of every element appearing in the viewport are reported by speech synthesis. The boundary location values correspond to tactile Braille rulers that have been taped on the screen frame. In this way, blind users can quickly and precisely obtain the general layout of the displayed document. In the active macro-analysis, the blind user explores the document by moving his or her finger on a touch sensitive screen. The type of the HTML element that is touched is represented by an auditory icon. Auditory icons were chosen as non-speech sounds since they facilitate an intuitive recognition of the related HTML element type. For example, typewriter and camera shutter sounds are employed to represent texts and images. The screen location of the element is provided by the spatial location of the auditory icon in a 3D sound space which represents the HTML document.

The micro-analysis phase allows blind users to retrieve the content of the HTML elements, that have been found during the macro-analysis exploration. If the user points on textual information, speech synthesis with tonal variations and simple sound effects (such as beeps), as employed by James (1996), presents the text itself as well as textual differentiation such as links or emphasis. If the user explores an image, the sound is related to the pictorial attributes of the touched pixel (for example colour or contour) and the spatial location of this sound in the virtual 3D audio space is related to the finger position within the image.

A first prototype using simple auditory icons has been implemented and evaluated. The results showed that the 3D auditory representation of the HTML document allows a good global layout determination and that the passive mode, in which the text-to-speech converter describes the screen's spatial organisation, is useful.

From the ideas of the AB-Web project, Petrucci, Harth, Roth, Assimacopoulos & Pun (2000) developed WebSound, a generic Web sonification tool which allows to easily create, test and validate new sonification models and combines the haptic sense with an audio output, and its application as a 3D audio augmented Internet browser.

James (1996) tested four interface styles for presenting HTML structures in audio: One speaker and minimal sound effects, one speaker and many sound effects, many speakers and minimal sound effects, many speakers and many sound effects. The sound effects in the interfaces were selected with the idea of auditory icons in mind. An effort was made to choose sounds that seemed intuitively related to the structural element they were meant to represent. If there was no obvious sound, a short abstract effect was used similar to earcons. His results showed for example that the use of three different speakers to present heading levels was confusing and that natural sounds are more distinguishable and easier to remember than abstract sounds.

Portigal (1994) used abstract sounds, similar to earcons, in his work on displaying document structure in sound. When a new section is opened, a cue indicates its location within the document. These sound cues consist of two sequences of tones. The number of tones in the first sequence indicates the number of chapters, with a pause after the current chapter. The number of tones in the second sequence indicated the number of subsections in the current chapter, with a pause after the current subsection. Each note is an eighth note, with the pause being a quarter note and the break between each sequence being a whole note. The first sequence began at middle C. The second sequence started one octave above the note that preceded the pause in the first sequence. His research failed to show any advantages from using unstructured bursts of sound.

2.3.7 Hearcons

Donker, Klante & Gorny (2002) proposed “Hearcons” which they define as three dimensional auditory objects which are positioned in an auditory interaction realm (AIR) to overcome the layout barrier of web pages. These are abstract synthesized sound similar to earcons. With a pointing device like a graphics tablet or a joystick the user can interact with these hearcons as a sighted user is interacting with his mouse in a graphical user interface. After a hearcon is selected by a click with the pointing device or by a spoken command an application function can be triggered, for example a screen reader reading the name of the hearcon.

A prototype of the proposed system, called „AIR-Client“, accesses the Document Object Model (DOM) of a standard web browser and transforms the content, structure and style of the document into an internal data structure. In the recent version of the AIR-Client the number of identified objects on a web page is restricted to consider only the most frequently appearing objects such as headings, paragraphs, images and links. For these objects they determined the attributes content, position, dimensions, type and, if available, the alternative text.

These four categories of layout objects are presented by the following motives: Headings are represented by horn timbres assigned with fanfare rhythms. Paragraphs are characterized through sounds with xylophone timbre playing the rhythm of a news ticker. For images a rhythm of Edward Grieg out of Morgenstimmung was used in some variations since this has a pictorially descriptive character. Links are represented by other synthesized timbres and corresponding rhythms.

As output devices they used both, headphones as well as an array of four plane loudspeakers which were put on a virtual wall in front of the user. Using headphones, the auditory user interface is mapped to the inner surface of a hemisphere section. The user interface has a horizontal extension (azimuth) of 180 degrees as well as a vertical extension (elevation) of 120 degrees and surrounds the visual field of the user in a distance of 3 meters. Using loudspeakers, the user interface is constituted by the position of the four loudspeakers in the corners of the user interface. The chosen experimental setup determined a horizontal extension (azimuth) of 82 degrees and a vertical extension (elavation) of 68 degrees. The smallest distance of the user to the user interface was 100 cm. As an input device a graphics tablet was used.

When the pointer approaches a layout object, its hearcon gets louder. When the mouse pointer is exactly located on a hearcon, it reaches its maximum volume. By the use of a suitable distance function they achieve to have only five to six perceivable hearcons. In addition to the volume alteration they also implemented a focussing function, which operates like a magnifying glass in combination with a fish eye view: Hearcons near the pointer explode and hearcons that are far away, near the edge of the display area, contract and sound like only one hearcon.

The evaluation results of a usability test with seven blind experts solving three tasks with the AIR-Client revealed that the invited experts were not able to process their tasks more effectively and more efficiently with the proposed system compared to their presently used screen reader. They needed considerably more time and identified less relevant layout objects on the web pages. Furthermore, they could not identify the rough structure of the layout. One hypothesis for this result is, that blind persons develop completely different mental models for the orientation in abstract material than sighted persons, who are dominantly relying on visual cues and develop spatial models. Blind users don't want to know what a web page exactly looks like, but they want to understand the object structure.

2.3.8 TactoWeb and MaskGen

Petit, Dufresne & Robert (2011) developed TactoWeb. This is a web browser allowing users with visual impairment to spatially explore web pages using tactile and auditory feedback. This exploration is made to work with the Tactograph device or the Logitech [iFeel mouse]. With TactoWeb, they would like to prove that a spatial exploration with tactile and audio feedback is more effective, more efficient and more satisfying than a sequential exploration with only audio feedback used by tradition tools giving access to web pages such as screen readers.

The Tactograph device enables tactile exploration of a two-dimensional 27.9 x 21.6 cm surface. Tactile feedback is produced by a Latero tactile cell, originally named STReSS (Stimulator of Tactile Receptors by Skin Stretch) 2 (Wang & Hayward, 2006). This tactile cell is really different from other tactile devices because it stimulates the tip of our finger by laterally stretching and contracting the skin. It allows a much finer touch rendering than others devices that usually use actuators going vertically in and out, such as Braille cells. The surface of the tactile cell is 9 x 11 mm, approximately matching the skin surface on the tip of our index finger. To laterally deform the skin, 64 actuators (8 x 8) move from right to left producing tactile sensations like vibration, undulations, etc. The distance between each actuator is 1.2 × 1.4 mm and each of them can bend from 0.1 mm to the left or to the right. The 64 actuators can be controlled to generate fine static or dynamic tactile rendering. Static rendering includes vibrations, fixed unilateral and radial undulations, vectorial shapes and points. Dynamic rendering includes moving unilateral or radial undulations, various tactile rendering depending on the speed of movement, and various tactile rendering depending on the direction of motion.

The Latero is installed inside a shell that looks like a computer mouse with two buttons on each side of the mouse. These buttons will allow the user to perform quick actions such as select a link, make the system repeat the audio feedback or going back in the Web browser history. Inside its shell, the Latero is mounted on a 2D exploration surface, being attached by a three joints arm. So with the Tactograph, it is possible to explore an area, while having tactile and audio feedbacks adapted to the location where the Latero is located on the exploration surface.

TactoWeb is able to adapt the following HTML elements: Headings, links, paragraphs, menus, unnumbered and numbered lists, image-maps, groups of buttons and form elements. Tactile rendering must be correctly chosen to fit each HTML element. Static tactile renderings are useful for basic features. For example, each heading of the same level uses the same static tactile rendering. Dynamic tactile rendering, on the other hand, can serve as a guide for the user. For example, when errors occur after validating a form, dynamic undulations can be used to guide the user towards the fields responsible.

TactoWeb gets Document Object Model (DOM) tree as well as the CSS tree from the URL entered by the user. The CSS tree is used in order to identify each formatting section of the web page and the layout settings of the different HTML elements like position, length, height, etc. Finally TactoWeb displays all the elements through the Tactograph device and the speech synthesis is used as a legend for the web page. It speaks all the text content of the Web page like paragraphs, headings, labels, etc.

They also plan TactoWeb to supporting the Logiteh [iFeel mouse], although the iFeel mouse does not have as fine tactile rendering as the Tactograph, because actually, the Tactograph is still pretty expensive.

Petit, Dufresne, Levesque, Hayward & Trudeau (2008) also developed another application for the Tactograph device called MaskGen, which helps to make schoolbook illustrations accessible for students with visual impairment. MaskGen allows the user to explore these illustrations giving a good idea of their structure thanks to tactile and audio feedbacks. This is exactly what they want to achieve for web pages with TactoWeb.

However, this is early work and they did not describe any experimentation conducted. As no evaluation has been carried out, the success of this system cannot be gauged.

2.3.9 Auditory Maps

Blattner, Papp & Glinert (1992) added earcons to two-dimensional maps. They used a map of floor plans of the Lawrence Livermore National Laboratory. To this they added information in sound such as type of computer equipment contained in the building, security clearance required and jobs of those in the building. Hierarchical earcons were used for the sounds. For example, a three note saxophone earcon indicated an administrative building and a tom-tom represented the security restriction. The faster the tom-tom was played the higher the restriction. The user could click on a building to hear what it contained or they could use an area selector. Any building which was within the area selector would play its earcons, concurrently with the other buildings. This technique allowed much more data to be represented than in the graphical case. Although no experimental testing was reported by Blattner et al. the system seemed to have much potential. A similar approach was also taken by Kramer (1992).

2.3.10 Algebra Earcons

Stevens, Brewster, Wright & Edwards (1994) used earcons to provide a method for blind readers to glance at algebra. They say that a glance, or overview, is very important for planning the reading process. There is currently no way to do this. The parameters for manipulating earcons are very similar to those describing prosody in speech. Stevens et al. combined algebra syntax, algebraic prosody and earcons to produce a system of algebra earcons. They describe a set of rules that can be used to construct algebra earcons from an algebra expression. Different items within an algebra expression are replaced by sounds with different timbres, such as: Piano for base-level operands, violin for superscripts and drums for equals. Rhythm, pitch and intensity are then defined according to other rules. Stevens et al. experimentally tested the earcons to see if listeners could extract algebraic structure from the sounds and identify the expressions they represented. Their results showed that subjects performed significantly better than chance. This work again shows that earcons are a flexible and useful method of presenting complex information at the user interface.

2.3.11 Haptic Emoticons

Rovers & Van Essen (2004) proposed the use of haptic icons similar to tactons as an alternative to emoticons in instant messaging applications. An emoticon, also called a smiley, is a sequence of ordinary printable characters or a small image, intended to represent a human facial expression and convey an emotion. A haptic emoticon is a vibrotactile sensation that can be used to express emotions in place of emoticons. Rovers & Van Essen (2004) suggest that these haptic icons could be triggered by typing the textual equivalent (just like emoticons), or via special input devices (for example by making a gesture on a touchpad). They would then be sent to a vibrating device at the receiver end which could be a vibration enabled mouse or joystick, or a small vibrotactile actuator attached to the receiver’s clothing or inside a shoe (Rovers & Van Essen, 2005). The proposed haptic icons would be designed using parameters such as frequency, amplitude and duration, and would use metaphorical mappings that were related to what they represented. No evaluation of these haptic emoticons has been carried out, but this proposal indicates that there is interest in the area of haptic icons for communication.

Chan (2004) also created a set of haptic icons or touch icons for use in a collaborative application to communicate information such as whether the user is currently in control of the application, waiting for control or observing, and if anyone else has requested control. These haptic icons were presented via the Logitech [ifeel mouse]. The parameters of vibration which can be manipulated on the ifeel mouse are frequency, amplitude, and rhythm, so these parameters were used to create a set of seven haptic icons. An evaluation by Chan, MacLean & McGrenere (2005) of these seven haptic icons under various degrees of workload (when the user was also engaged in a visual or auditory task) showed identification rates of 95 per cent on average, with no significant differences in performance with or without distractor tasks, indicating that vibrotactile icons can be an extremely effective means of communicating information.

2.3.12 Vibrocons

Van Erp & Van Veen (2001) designed “Vibrocons”, which are vibrotactile icons similar to tactons for an in-car navigation system, using vibrotactile devices mounted in a car seat. The vibrocons encoded the direction the driver should turn using spatial location, with a vibration on the left leg indicating a left turn, and a vibration on the right indicating that the driver should turn right. The distance to the turning was encoded in the temporal pattern, with the direction information presented more frequently as the turning point approached. These vibrocons are designed according to the guidelines for the use of vibrotactile displays in human computer interaction of Van Erp (2002) stating that messages should be self-explaining. The results of this study indicate that vibrotactile feedback can be effective in improving reaction times when people are under workload, and that more than one dimension of information can be interpreted from vibrotactile stimuli.

Van Veen & Van Erp (2000) also investigated the use of a vibrotactile vest to present information to fighter jet pilots. Tactile display are particularly useful in this setting as vision can be affected when pilots are in g-load conditions. The tactile vest consists of

128 tactile actuators distributed across the torso. In this display the spatial location of the activated tactile actuators can be used to present a variety of information. For example, activating a single point could indicate the direction of a target or destination. In addition, a line of actuators could be activated to indicate an artificial horizon. These uses are examples of a direct approach similar to Vibrocons described above. The tactile display could also be used to present coded information such as fuel supply, speed, altitude or time, or to enable discreet communication between pilots or crew members. A study was conducted where users were asked to identify whether a vibration had occurred on their left or their right hand side, both in normal conditions and under g-load. The results showed that performance was consistently high with 85 to 100 per cent correct, and that vibrotactile perception was not significantly impaired under g-load.

Gemperle, Ota & Siewiorek (2001) also designed a wearable tactile vest for tactile information display, the main application of which was the display of navigation information. This vest could present information on direction, rotation, speed and acceleration to a range of users including pedestrians and cyclists, as well as car drivers. The stimulation of a single actuator on the appropriate location could indicate direction, while a movement across several actuators could signal rotation. Changing the frequency, intensity or duration of the vibrations could indicate an increase or decrease in speed or acceleration. This system has not been evaluated beyond pilot testing, so it is not possible to assess how successful the method of information presentation is.

Ho, Spence & Tan (2005) and Ho, Tan & Spence (2005) investigated the use of vibrotactile cues to alert drivers to potentially dangerous events on the road. Their results showed that drivers responded significantly faster to alerts provided via vibrotactile stimuli than those presented by either audio or visual stimuli and that more rapid responses were facilitated by presenting the vibrotactile stimulus to the side of the body (front or back of torso) that corresponded to the location of the event. These results indicate that vibrotactile messages can be attention grabbing, and that people are able to accurately localise vibrations presented to the front or back of the torso.

In addition to aiding navigation in vehicles, vibrotactile feedback has also been used to help blind or visually impaired persons to navigate. For example, Ross & Blasch (2000) designed a wearable tactile display, which indicated whether the user was walking in the right direction or if a change of direction was needed. Three vibrotactile actuators were located on the user’s upper back. If the user was on target, the centre actuator was activated every two second to indicate this. The left and right actuators were activated to indicate that the user was off target in that direction. Their results showed that the tactile display resulted in better performance and was preferred by users when compared with a speech interface as well as a non-speech audio interface.

In addition to providing navigation information, tactile displays have also been used to help blind persons with obstacle avoidance. For example the People Sensor developed by Ram & Sharf (1998) uses sensors to locate (and distinguish between) people and objects, and indicates their location to the user through vibrotactile output. Two different types of pager motors are attached to the user’s waist. As the motors are different types, they produce different sensations so one is used to represent people and the other to represent inanimate objects. The distance to the person or object is indicated to the user by the rate at which vibrotactile pulses are presented, with the pulses becoming more frequent as the person or object gets closer. As no evaluation has been carried out, the success of this system cannot be gauged.

2.3.13 CrossTrainer

From the ideas of crossmodal icons, Hoggan & Brewster (2010) developed the CrossTrainer, which is a mobile game with crossmodal audio and tactile feedback used in an exploratory 8-day field study focusing on the longitudinal effects on performance with audio and tactile feedback, the impact of context such as location and situation on performance and personal modality preference. The results of this study indicate that crossmodal feedback can aid users in entering answers quickly and accurately using a variety of different widgets. This study also shows that there are times when audio is more appropriate than tactile and vice versa and for this reason devices should support both tactile and audio feedback to cover the widest range of environments, user preference, locations and tasks.

2.4 Accessibility APIs

The Accessibility Application Programming Interface (API) of the operating system, acts as the global interface between user agents, authoring tools and the assistive technologies running on a system. It enables assistive technologies to dynamically access and update the structure, content and style of a structured document by proving the relevant information and control.

In modern accessibility APIs, user interfaces are represented as a hierarchical tree. For example, an application window would contain several objects, the first of which might be a menu bar. The menu bar would contain a number of menus, each of which contains a number of menu items, and so on. If a structured document is contained in a user interface, for example within a web-browser application window, the Document Object Model (DOM) is used for representing that content.

2.4.1 Document Object Model (DOM)

The Document Object Model (DOM) 4.0, developed by Van Kesteren, Aryeh, Russell & Berjon (2015), is a platform- and language-neutral interface that allows programs and scripts to dynamically access and update the content, structure and style of structured documents.
The Document Object Model (DOM) is an application programming interface (API) for valid and well-formed documents. It defines the logical structure of documents and the way a document is accessed and manipulated. With the Document Object Model, programmers can build documents, navigate their structure, and add, modify, or delete elements and content. Anything found in a document can be accessed, changed, deleted, or added using the Document Object Model. As a W3C specification, one important objective for the Document Object Model is to provide a standard programming interface that can be used in a wide variety of environments and applications. The DOM is designed to be used with any programming language.

In the DOM specification, the term document is used in the broad sense. Is is being used as a way of representing many different kinds of information that may be stored in diverse systems, and much of this would traditionally be seen as data rather than as documents but the DOM may be used to manage this data.

The DOM is based on an object structure that closely resembles the structure of the documents it models. In the DOM, documents have a logical structure which is very much like a tree. The DOM presents documents as a hierarchy of Node objects that also implement other, more specialized interfaces. Some types of nodes may have child nodes of various types, and others are leaf nodes that cannot have anything below them in the document structure. A child is an immediate descendant node of a node.

The Node interface is the primary datatype for the entire Document Object Model. It represents a single node in the document tree. The most important subtypes of nodes are document, element and text.

The Document node represents the entire document. Conceptually, it is the root of the document tree, and provides the primary access to the document's data. The root node is a node that is not a child of any other node. All other nodes are children or other descendants of the root node. Element and Text nodes cannot exist outside the context of a Document.

Each document contains one or more elements. The Element node represents an element in a document. Each element has a type, identified by name, and may have a set of attributes. Each attribute has a name and a value.

The Text interface inherits from CharacterData and represents the textual content of an Element or attribute. If there are no other elements inside an element's content, the text is contained in a single object implementing the Text interface that is the only child of the element. If there are other elements contained, these element and text node items form a list of children of the element. Two nodes are siblings if they have the same parent node.

2.4.2 IAccessible2

Microsoft has detected the problem of user interface accessibility as early as 1997, when the Microsoft Active Accessibility API MSAA (Microsoft, 1997) had been specified. With today's rich user interfaces including many new controls and features, this API comes to border. With the introduction of Windows Vista the User Interface Automation API (UIA) (Microsoft, 2005) has been introduced. As an alternative the IAccessible2 API (ISO/IEC 13066, 2007) as a platform independent standard accessibility API is used under Windows too.

In the Linux domain, as a toolkit-neutral way of providing accessibility facilities in applications the Assistive Technology Service Provider Interface (AT-SPI) (Gnome, 2002) was developed by the Gnome Accessibility Project in 2002 for providing a logical representation of the content of the application. AT-SPI can also be used for automated testing of user interfaces with tools such as Linux Desktop Testing Project and Dogtail. In the future, the IAccessible2 API (ISO/IEC 13066, 2007) as a cross-platform standard will be used.

2.4.3 iOS UIAccessibility Protocol

Since version 3.0 of Apple's iOS operating system released in 2009, the UIAccessibility API (Apple, 2009) informal protocol provides accessibility information about an application's user interface elements. Assistive technologies, such as [VoiceOver], convey this information to users with disabilities to help them use the application. This API can be used by App developers to provide information of their application to the assistive technology. But at the moment it is on the other hand not possible to developing an assistive technology accessing this API. Therefore VoiceOver it the one and only assistive technology within iOS.

2.4.4 Android Access Framework

Since the release of Google Android version 1.6 in 2009, the Android Access Framework (Android, 2009) is used to represent screen content and changes to it as well as for querying the global accessibility state of the system. In contrast to Apple's iOS, the Android Access Framework allows and welcomes the development of user specified assistive technologies (called Accessibility Services) to serve all the special needs.

Android does also provide a set of preinstalled assistive technologies like the screen reader [TalkBack].

2.5 Content

There are many different content types of digital structured documents available on the Internet, for example text-documents, spreadsheets, presentations, drawings, mathematics or web-sites and manifold different formats are used to store this content and exchanging it over the Internet.

The Web Content Accessibility Guidelines (WCAG) 2.1 (Kirkpatrick, O'Connor, Campbell & Cooper, 2015) covers a wide range of technology independent recommendations for making structured documents more accessible for a wide range of persons with disabilities. Following these guidelines will also often make the documents more usable to users without disability in general. In addition, the Techniques and Failures for Web Content Accessibility Guidelines 2.0 (Cooper, Kirkpatrick & O'Connor, 2016) guides authors and evaluators on meeting the WCAG success criteria by providing general as well as technology specific techniques and failures.

The User Agent Accessibility Guidelines (UAAG) 2.0 (Allan, Lowney, Patch & Spellman, 2015) guides developers in designing user agents that will improve accessibility through their own user interface and their ability to communicate with assistive technologies.

The Authoring Tool Accessibility Guidelines (ATAG) 2.0 (Richards, Spellman, Treviranus & May, 2015) provides guidelines for designing authoring tools that are both more accessible to authors with disabilities and designed to enable, support, and promote the production of more accessible content by all authors.

Some work in this area has been done by Davishy, Hutter, Horvath & Dorigo (2010), Darvishy, Leemann & Hutter (2012), Bianchetti, Erle & Hofer (2012), Darvishy & Hutter (2013), Darvishy, Nevill & Hutter (2016) and Darvishy (2018), who developed plug-ins for authoring tools to assist authors in creating accessible content through prompts, alerts, checking and repair functions.

In this chapter, the different content formats for storing and exchanging digital structured documents over the Internet are described and their accessibility features are discussed.

2.5.1 Hypertext Markup Language (HTML)

The Hypertext Markup Language (HTML) is the publishing language of the World Wide Web. In Hypertext a structured document is splitted into multiple small semantically self-contained documents. These are linked together using Hyperlinks, not in a sequential structure, but in an arbitrary structure of a web according to their semantic relationship.

HTML 4.0 (Raggett, Le Hors & Jacobs, 1999) provides structure tags for basic structure elements of web-sites like headings, lists, or tables. It does not provide native support for rich components, for example sliders or trees used in modern rich internet applications.

For that purpose the Accessible Rich Internet Applications (WAI-ARIA) 1.1 standard has been developed by Diggs, McCarron, Cooper, Schwerdtfeger & Craig (2017), as an extension for the HTML 4.0 standard. Accessibility of web content requires semantic information about widgets, structures, and behaviours, in order to allow assistive technologies to convey appropriate information to persons with disabilities. This specification provides an ontology of roles, states, and properties that define accessible user interface elements and can be used to improve the accessibility and interoperability of web content and applications. These semantics are designed to allow an author to properly convey user interface behaviours and structural information to assistive technologies in document-level markup.

The emerging HTML 5.2 standard (Faulkner, Eicholz, Leithead, Danilo & Moon, 2017), the 5th major revision of the core language of the World Wide Web, has now integrated native support for rich components. Therefore the WAI-ARIA extension is no longer required since HTML 5.0. In this version, new features are introduced to help Web application authors, new elements are introduced based on research into prevailing authoring practices, and special attention has been given to defining clear conformance criteria for user agents in an effort to improve interoperability.

2.5.2 Portable Document Format (PDF)

The Portable Document Format (PDF) specifies a digital format for representing structured documents to enable users to exchange and view structured documents independent of the environment in which they were created or the environment in which they are viewed or printed.

In first versions of PDF the logical document structure was not accessible because only the visual physical structure had been stored within a PDF file similar to PostScript.

Since PDF 1.7 (ISO 32000, 2008) became an international ISO standard, accessibility tags had been added for providing the logical document structure within a PDF file: In additional to the content stream, the accessibility tag tree is stored in a separate stream, where each object within the content stream may be linked to an object within the structure tag stream. This enables a 1:1 relationship between the logical and the physical document structure. However, there was no standardisation about the tag names. This leaded to the confusing situation that different authoring tools were using different tag names for one and the same structure element.

Since 2012 PDF Univarsal Access (PDF/UA) (ISO 14289, 2012) is the international standard for accessible PDF technology adding many benefits like reliable text re-flow on small screens or powerful navigation options.

2.5.3 Open XML Paper Specification (XPS)

The Open XML Paper Specification (XPS) also referred to as OpenXPS is an open specification for a page description language and a fixed-document format. Microsoft developed it as the XML Paper Specification (XPS). In June 2009, it was adopted as an international standard (ECMA 388, 2009). Since Windows 8 XPS was replaced with the Open XML Paper Specification (OXPS) format which is not natively supported in older Windows versions.

The XPS document format consists of structured XML markup that defines the physical layout of a document and the visual appearance of each page, along with rules for distributing, archiving, rendering, processing and printing the documents. XPS is a page description language which can describe a single page or a document containing multiple pages. The description includes all the text and graphics that appear on a page. Like other page description languages such as PDF, these physical structure elements are defined independently of a particular operating system, printer or viewing application. Therefore, the document's appearance is consistent regardless of the specific printer or viewer used.

In XPS there are no accessibility tags. Document content is simply marked as particular logical document structure elements instead. As with PDF, this enables a 1:1 relationship between the logical document structure and the physical document structure.

2.5.4 Office Open XML (OOXML)

The Office Open XML Document File Formats (OOXML), also known as OpenXML, is an XML-based file format for storing different types of logical structured documents such as text-documents, spreadsheets, presentations and graphics. It was developed with the aim of providing an open, XML-based file format specification for office applications.

The specification was developed by Microsoft as successor of the depreciated proprietary vendor specific binary legacy document file formats such as DOC, XLS and PPT. The format was initially standardised by ECMA International as ECMA 376 (2006) in 2006 and has been adopted by the ISO and IEC as ISO/IEC 29500 (2008) in later versions. A second version was released in December, 2008, and a third version of the standard released in June, 2011. Starting with Microsoft Office 2007, the Office Open XML file formats have become the default target file format of Microsoft Office.

Its publication benefits organizations that intend to implement applications capable of using the format, commercial and governmental entities that procure such software, and educators or authors who teach the format. Ultimately, all users enjoy the benefits of an XML standard for their documents, including stability, preservation, interoperability, and ongoing evolution.

2.5.5 Open Document Format (ODF)

The Open Document Format for Office Applications (ODF), also known as OpenDocument, is an XML-based file format for storing different types of logical structured documents such as text-documents, spreadsheets, presentations, graphics and mathematical formulae and equations. It was developed with the aim of providing an open, XML-based file format specification for office applications.

The standard was developed by a technical committee in the Organization for the Advancement of Structured Information Standards (OASIS) consortium. It was based on the Sun Microsystems specification for OpenOffice.org XML, the default format used by OpenOffice.org, which had been specifically intended to provide an open standard for office documents. In addition to being an OASIS standard, it was published as an ISO/IEC international standard (ISO/IEC 26300, 2006).

While the OpenDocument specification was going through an extensive accessibility review, many of the components it is built on, for example SMIL for audio and multimedia and SVG for vector graphics, have already gone through the World Wide Web Consortium (W3C) Web Accessibility Initiative processes. The OASIS OpenDocument technical committee released a draft of OpenDocument 1.1 in 2006. This is a very minor update to the specification to add accessibility information, mainly for soft page break, table headers, presentation navigation, alternative text and captions, and specifically stating that spreadsheets may be embedded in presentations.

2.6 Conclusions

Literature within related fields of research about what structured documents are in general was reviewed and the essential components involved in the process of non-visual reading, creating and editing of structured documents had been identified. The different approaches available for each component were described in greater detail and their potential applications in a user interface for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices had been discussed. In addition, some examples of existing systems that use each non-visual approach in their human computer user interfaces were provided. The investigation of these general research issues about the reading of structured documents for blind and visually impaired persons leaded to the following research questions for initial research.

On its way from the author to the reader, a digital structured document passes several components. These components have to work together to provide full access for the user. If an accessibility feature is not implemented in one component, it does not result in an accessible user experience. An interesting research question for initial research regarding satisfaction would be: How satisfied are blind and visually impaired persons with their current general situation as to the reading of structured documents ?

There are many different content types of digital structured documents available on the Internet, for example text-documents, spreadsheets, presentations, drawings, mathematics or web-sites and manifold different formats are used to store this content and exchanging it over the Internet. Interesting research questions for initial research regarding content would be: On which different physical media are structured documents provided to blind and visually impaired persons ? How often do the persons concerned receive structured document on each physical medium ? And on which preferred physical medium would they wish to have their structured documents ? In which different digital formats are structured documents provided to blind and visually impaired persons ? How often do the persons concerned receive structured documents in each digital format ? And in which preferred digital format would they wish to have their structured documents ? What logical structure elements do blind and visually impaired persons use for the navigation within structured documents ? How important is each of this logical structure elements for them ? Which accessibility problems do blind and visually impaired persons have as to the reading of structured documents ? How often do each of these accessibility problems occur to them ? And if blind and visually impaired persons are unable to read a specific structured document or parts of it because the document is inaccessible, what accessibility solutions do they have to solve this problem at the moment ?

The user agent or authoring tool parses this content to read a structured document or transforms the structured document to store it in a specific content format and enables assistive technologies to dynamically access and update the structure, content and style of the document by proving the relevant information and control via the accessibility API of the operating system.

The accessibility Application Programming Interface (API) of the operating system, acts as the global interface between user agents, authoring tools and the assistive technologies running on a system. It enables assistive technologies to dynamically access and update the structure, content and style of a structured document by proving the relevant information and control. In modern accessibility APIs, user interfaces are represented as a hierarchical tree. If a structured document is contained in a user interface, for example within a web-browser application window, the Document Object Model (DOM) is used for representing that content.

The assistive technology, dynamically accesses and updates the structure, content and style of the structured document provided by the accessibility API of the operating system and provides an alternative user interface for presentation, navigation and manipulation to the user by employing various different modalities. In addition to traditional screen readers there are many other research projects aiming to provide a solution for non-visual presentation, navigation and manipulation of different types of structured data. Interesting research questions for further research regarding assistive technologies would be: Which types of assistive technologies do blind and visually impaired persons use for the reading of structured documents for desktop computers ?

Which types of assistive technologies do the persons concerned use for the reading of structured documents on mobile and wearable devices ? Which products for each type of assistive technologies do they use ? And finally, what features would blind and visually impaired persons wish to have for a novel concept of an assistive technology in support of them for the reading of structured documents ?

The modality presents particular pieces of information to the user in a modality specific way by stimulation of one or more human sense using specific hardware devices. The different approaches available can be grouped into two main categories: Text-based and iconic modalities. In the visual domain there is graphical text and its iconic counterparts are graphical icons. In the auditory domain there is synthetic speech as text-based modality and there are auditory icons, earcons and spearcons as iconic counterparts. In the tactile domain one common form of text-based output is Braille in addition to Vibratese and the iconic counterparts are tactons and hapticons. The concept of crossmodal icons had been introduced which are abstract icons that can be automatically instantiated as either an earcon or tacton. Interesting research questions for further research regarding modalities would be: Which input modalities do blind and visually impaired persons use for the reading and navigation within structured documents on desktop computers ? Which input modalities do they use for the reading and navigation within structured documents on mobile and wearable devices ? Which hardware devices are blind and visually impaired persons using for the reading of structured documents at the moment ? Which hardware devices are they planning to purchase in the future within the next 2 to 3 years ?

The reader (a persons who reads a structured document) or the author (a person who creates and edits structured documents) knows how to use the assistive technologies, user agents or authoring tools as well as the different modalities employed for presentation, navigation and manipulation to read, create and edit structured documents. Interesting research questions for initial research regarding readers and authors would be: To which visual impairment category, according to the WHO International Classification of Diseases ICD (WHO, 1994), do the research subjects belong to ? Do the persons concerned also have a visual field restriction in addition the restricted visus ? Are they visually impaired by birth ? What computer affinity do blind and visually impaired persons come with ? In what occupational situations are blind and visually impaired persons ? What educational background do blind and visually impaired persons have Are blind and visually impaired persons interested in being involved in further research by participating in a pilot test or by testing of a prototype of a novel concept for an assistive technology for the reading of structured documents?

During the processing by the different components for document development and document recognition, structured documents take on different forms. The logical document is formed by the logical document structure together with the document content. The document content describes to information contained within a structured document in a medium and presentation independent way. The logical document structure defines how the content of a structured document is semantically organised in a medium and presentation independent way. The physical document is formed by the physical document structure together with the document content. The physical structure of a document specifies how the document content is physically organised on a specific medium. The digital logical document image is the rendered or sampled form of the physical document in a hardware independent way. The analogue physical document image is created by presenting the digital logical document image to the user on a specific hardware output device.

3 Research Methodology

3.1 Introduction

The user interface has been developed by employing the following social research methods using a natural science epistemological model (positivism), an objective ontology (objectivism), a quantitative research strategy and an iterative research approach. Figure 3.1 gives an overview over this research methodology:

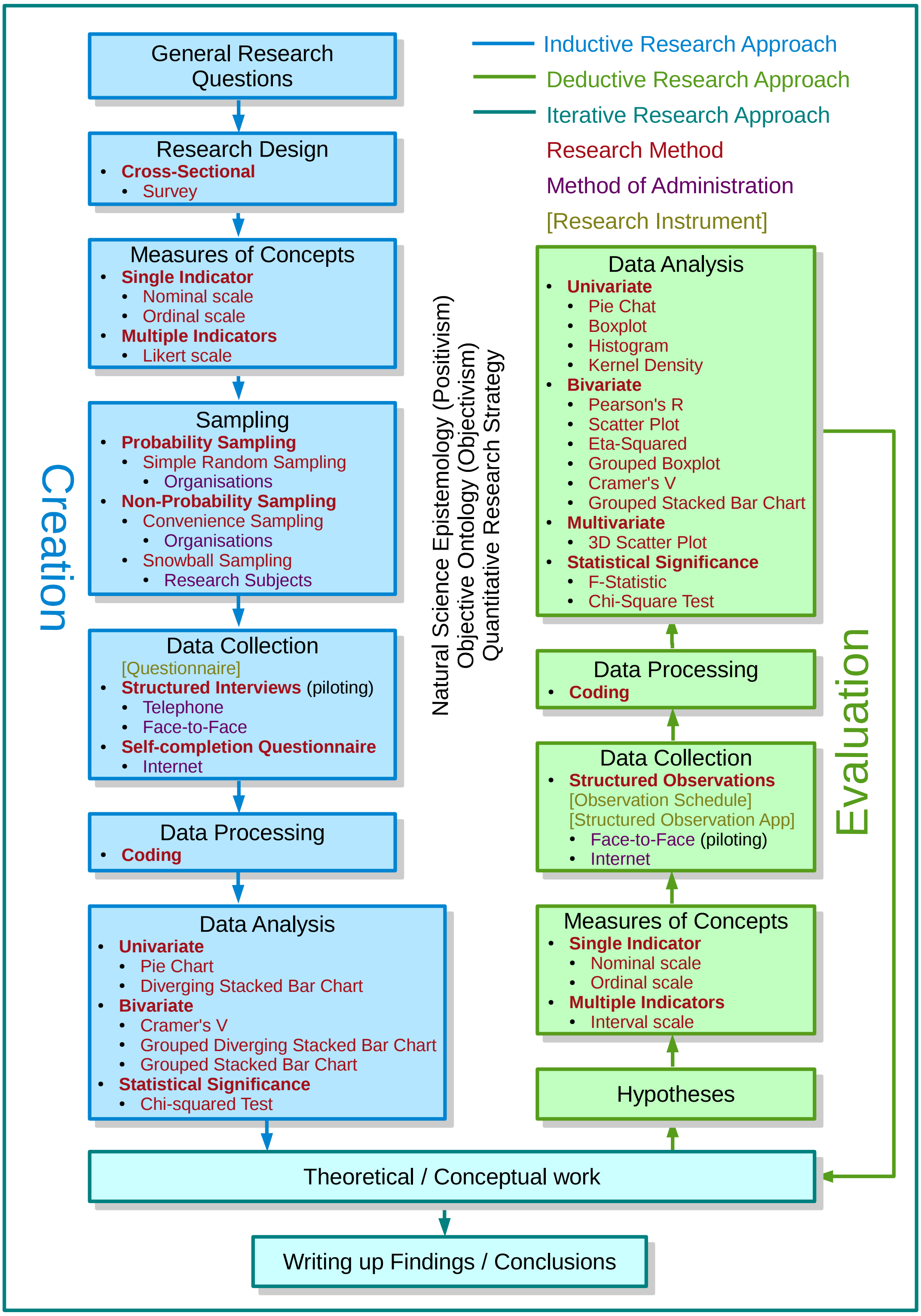


Figure 3.1: Research Methodology Overview

3.1.1 Research Strategy

According to Bryman (2016) the research strategy characterises the nature of the link between theory and research. There is the matter of whether data are collected to build or to test theories. This relationship between theory and research is important because it provides a backcloth and rationale for the research that is being conducted. It also provides a framework within which social phenomena can be understood and the research findings can be interpreted.

The user interface as described in Chapter 5 has been created using an inductive research approach where general research questions stand at the beginning as the hypotheses and the theory, in this case the user interface design Chapter 5.2 and the user interaction design Chapter 5.3, is the outcome of the research. Knowledge is arrived at through the gathering of facts that provide the basis for laws. In other words, the process of induction involves drawing generalizable inferences out of observations. This is the principle of inductivism. Hence data are collected to build the theories.

Once this phase of theoretical reflection on a set of data had been carried out, further data had been collected in order to establish the conditions in which the theory will and will not hold. For this evaluation of the user interface, a deductive research approach has been used. The purpose of theory is to generate the hypotheses that can be tested and that will thereby allow explanations of laws to be assessed. This is the principle of deductivism. Hence data are collected to test the theories, in this case the user interface design and the user interaction design. Deduction represents an alternative strategy for linking theory and research. Theory and the hypotheses deduced from it come first and drive the process of gathering data. Deductive theory represents the commonest view of the nature of relationship between theory and social research. On the basis of what is known about in a particular domain and of theoretical considerations in relation to the domain, hypotheses are deduced that must then be subjected to empirical scrutiny. Embedded within the hypothesis will be concepts that will need to be translated into researchable entities that are so called operational terms.

The last step involves a movement that is in the opposite direction from deduction, it involves induction, as the implications of the findings from the theory that prompted the whole exercise had been inferred. These findings are fed back into the stock of theory and the research findings associated with the domain of non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices. Such a general strategy is often called iterative. It entails elements of induction as well as a modicum of deduction and involves a weaving back and forth between data and theory, this is particularly evident in grounded theory.

Quantitative research can be constructed as a research strategy that emphasizes quantification in the collection and analysis of data. Quantitative researchers employ measurements.

The term empiricism is used in a number of different ways, but two stand out. First it is used to denote a general approach to the study of reality that suggests that only knowledge gained through experience and the senses is acceptable. In other words, this position means that ideas must be subjected to the rigours of testing before they can be considered knowledge. The second meaning of the term is related to this and refers to a belief that the accumulation of facts is a legitimate goal in its own right. It is the second meaning that is sometimes referred to as naive empiricism.

3.1.2 Epistemological Position

An epistemological issue concerns the questions of what is or should be regarded as acceptable knowledge in a discipline. A particularly central issue in this context is the question of whether the social world can and should be studied according to the same principles, procedures and ethos as the natural sciences.

In this research, a natural science epistemological model (positivism) has been employed because positivism is an epistemological position that advocates the application of the methods of the natural sciences to the study of social reality and beyond. Only phenomena and hence knowledge confirmed by the senses can genuinely be warranted as knowledge. This is the principle of phenomenalism. Positivism entails elements of both a deductive approach and an inductive research strategy. Therefore the role of research is to provide material for the development of laws as well as to test the theories.

3.1.3 Ontological Position

Questions of social ontology are concerned with the nature of social entities. Whether social entities can and should be considered objective entities that have a reality external to social actors, or whether they can and should be considered social constructions built up from the perceptions and actions of social actors. These positions are frequently referred to respectively as objectivism and constructionism.

An objective ontology has been employed in this research because science must and presumably can be conducted in a way that is value free, that is objective. Ojectivism is an ontological position that asserts that social phenomena confront us as external facts that are beyond our reach of influence and their meanings have an existence that is independent of social actors. It implies that social phenomena and the categories that we use in everyday discourse have an existence that is independent or separate from actors. There is a clear distinction between scientific statements and normative statements and a belief that the former are the true domain of the scientist. This last principle is implied by the first because the truth or otherwise of normative statements cannot be confirmed by the senses.

3.2 Creation

The user interface as described in Chapter 5 has been created using an inductive research approach where general research questions stand at the beginning as the hypotheses and the theory, in this case the user interface design Chapter 5.2 and the user interaction design Chapter 5.3, is the outcome of the research. Knowledge is arrived at through the gathering of facts that provide the basis for laws. In other words, the process of induction involves drawing generalizable inferences out of observations. This is the principle of inductivism. Hence data are collected to build the theories.

3.2.1 General Research Questions

Initially the research began with general research issues about the reading of structured documents for blind and visually impaired persons that needed to be investigated. These had been gradually narrowed down, so that they became research questions which take the form of hypotheses. The movement from research issues to research questions was the result of reading the literature relating to the issues, such as relevant theories and evidence as outlined in the literature review in Chapter 2. Embedded within the hypotheses are concepts that had been translated into researchable entities which are so called operational terms.

In this research, the following 10 concepts and their research questions about the attitudes of blind and visually impaired persons in relation to the reading of structured documents had been subjected Table 3.1:

Concept	Research Questions
General Satisfaction	• How satisfied are blind and visually impaired persons with their current general situation as to the reading of structured documents ?
Physical Media	• On which different physical media are structured documents provided to blind and visually impaired persons ? • How often do the persons concerned receive structured document on each physical medium ? • On which preferred physical medium would they wish to have their structured documents ?
Digital Formats	• In which different digital formats are structured documents provided to blind and visually impaired persons ? • How often do the persons concerned receive structured documents in each digital format ? • In which preferred digital format would they wish to have their structured documents ?
Assistive Technologies	• Which types of assistive technologies do blind and visually impaired persons use for the reading of structured documents for desktop computers ? • Which types of assistive technologies do the persons concerned use for the reading of structured documents on mobile and wearable devices ? • Which products for each type of assistive technologies do they use ?
Input Modalities	• Which input modalities do blind and visually impaired persons use for the reading and navigation within structured documents on desktop computers ? • Which input modalities do they use for the reading and navigation within structured documents on mobile and wearable devices ?
Structure Elements	• What logical structure elements do blind and visually impaired persons use for the navigation within structured documents ? • How important is each of this logical structure elements for them ?

Accessibility Problems	• Which accessibility problems do blind and visually impaired persons have as to the reading of structured documents ? • How often do each of these accessibility problems occur to them ?
Accessibility Solutions	• If blind and visually impaired persons are unable to read a specific structured document or parts of it because the document is inaccessible, what accessibility solutions do they have to solve this problem at the moment ?
Hardware Devices	• Which hardware devices are blind and visually impaired persons using for the reading of structured documents at the moment ? • Which hardware devices are they planning to purchase in the future within the next 2 to 3 years ?
Novel Features	• What features would blind and visually impaired persons wish to have for a novel concept of an assistive technology in support of them for the reading of structured documents ?

Table 3.1: Concepts and Research Questions about Attitudes

In addition to the attitudinal questions as described above the following 5 personal factual questions, in which the research subjects were asked to provide personal information about themselves, and question about their knowledge in this area, had been subjected Table 3.2:

Concept	Research Questions
Visual Performance	• To which visual impairment category, according to the WHO International Classification of Diseases ICD (WHO, 1994), do the research subjects belong to ? • Do the persons concerned also have a visual field restriction in addition the restricted visus ? • Are they visually impaired by birth ?
Computer Affinity	• What computer affinity do blind and visually impaired persons come with ?
Occupational Situation	• In what occupational situations are blind and visually impaired persons ?
Highest Education	• What educational background do blind and visually impaired persons have ?

Participation in Further Research	• Are blind and visually impaired persons interested in being involved in further research by participating in a pilot test or by testing of a prototype of a novel concept for an assistive technology ?

Table 3.2: Personal Factual and Knowledge Concepts and Research Questions

3.2.2 Research Design

The research design provides the framework for the collection and analysis of data. A choice of research design reflects decisions about the priority given to a range of dimensions of the research process. These include the importance attached to: expressing causal connections between variables, generalising to larger groups of individuals than those actually forming part of the investigation, understanding behaviour and the meaning of the behaviour in its specific social context.

In this research, a cross-sectional survey research design had been employed because more than one case had been examined at a single point in time using quantitative or quantifiable data with more than one variable in order to detect variation between cases and patterns of association between variables. The imanipulable variables were not manipulated during the research because the research had been conducted at a single point in time. Having a temporal appreciation of social phenomena and their interconnections.

More than one case had been examined because we are interested in variation. Variation can be established only when more than one case is being examined. For this research, a lot more than two cases had been selected for a variety of reasons: It is more likely to encounter variation in all the variables in which we are interested, finer distinctions between cases can be made and the requirements of sampling procedure as outlined in Chapter 3.2.4 are likely to necessitate larger numbers.

In cross-sectional design research, data on the variables of interest are collected more or less simultaneously at a single point in time. For example when an individual completes a questionnaire the answers are supplied at essentially the same time.

Quantitative or quantifiable data with more than one variable is required because in order to establish variation between cases and then to examine associations between variables, it is necessary to have a systematic and standardised method for gauging variation. One of the most important advantages of quantification is that it provides the researcher with a consistent benchmark.

With a cross-sectional design it is possible to examine relationships and patterns of association only between variables. There is no time ordering to the variables, because the data on them are collected more or less simultaneously at a single point in time and the variables are not manipulated during the research. All that can be said is that variables are related. This is not to say that it is not possible to draw ambiguity about the direction of causal inferences from research based on a cross-sectional design. Some of the variables in which social scientists are interested are often viewed as potentially significant independent variables, simply because they cannot be manipulated, other than by extreme measures.

Three of the most prominent criteria for the evaluation of social research are reliability, replication and validity. Reliability is concerned with the question whether the results of a study are repeatable. The term is commonly used in relation to the question of whether the measures that are devised for concepts in the social sciences are consistent. In order for replication to take place, a study must be capable of replication, it must be replicable. A further and in many ways the most important criterion of research is validity. Validity is concerned with the integrity of the conclusions that are generated from a piece of research. Measurement validity applies primarily to quantitative research and to the search for measures of social scientific concepts. Essentially it is to do with the question of whether a measure that is devised of a concept really does reflect the concept that it is supposed to be denoting. Internal validity relates mainly to the issue of causability. Internal validity is concerned with the question whether a conclusion that incorporates a causal relationship between two or more variables holds water. In discussion issues of causability, it is common to refer to the factor that has a causal impact as the independent variable and the effect as the dependent variable. External validity is concerned with the question of whether the results can be generalised beyond the specific research context.

Replicability is likely to be present in most cross-sectional research to the degree that the procedures are spelled out for. Internal validity is typically weak because it is difficult to establish causal direction from the resulting data. Cross-sectional research designs produce associations rather than findings from which causal inferences can be unambiguously made. External validity is strong when the sample from which data are collected has been randomly selected. When non-random methods of sampling are employed, external validity becomes questionable. The issues of reliability and measurement validity are primarily matters relating to the quality of the measures that are employed to tap the concepts in which the researcher is interested, rather than matters to do with the research design.

3.2.3 Measures of Concepts

Measures allow us to delineate fine differences between people in terms of the characteristic in question. Measurements give us a consistent device or yardstick for making such distinctions. A measurement device provides a consistent instrument for gauging difference. Measurements provide the basis for more precise estimates of the degree of relationship between concepts for example through correlation analysis.

In this research, the 10 concepts and their research questions about the attitudes of blind and visually impaired persons in relation to the reading of structured documents as well as the 5 personal factual concepts and research questions and questions about their knowledge in this area described in detail in Chapter 3.2.1 are measured as follows:

How satisfied blind and visually impaired persons are with their current general situation as to the reading of structured documents is measured by an ordinal scale variable with the categories unsatisfied, adequate and very satisfied. This scale is ordinal because it can be rank ordered according to their satisfaction where unsatisfied is less satisfied then very satisfied.

On which different physical media structured documents are provided to blind and visually impaired persons and how often this occurs is measured by a Likert scale because it consists of multiple items. For each physical medium item, an ordinal variable with the 4 categories never, seldom, often and very often is employed. This scale is ordinal because these categories can be rank ordered according to their frequency where never is less frequent then very often. On which preferred physical medium they would wish to have their structured documents is measured by a single nominal scale variable. This scale is nominal because the different physical media cannot be rank ordered in any way.

The Likert scale is a multiple-item or multiple-indicator measure of a set of attitudes relating to a particular area. The goal is to measure intensity of feelings about the area in question. It comprises a series of statements (items) that focus on a certain issue. Each research subject is then asked to indicate his or her level of agreement with the statement. Sets of attitudes always needs to be measured by batteries of indirect indicators. In this research, always four-point scales with 4 categories in which there is no neutral point are employed because it gets a more clear result where the research subject is forced to take a position.

In which different digital formats structured documents are provided to blind and visually impaired persons and how often this occurs is measured by a Likert scale because it consists of multiple items. For each digital format item, an ordinal scale variable with the categories never, seldom, often and very often is employed. This scale is ordinal because these categories can be rank ordered according to their frequency where never is less frequent then very often. On which preferred digital format they would wish to have their structured documents is measured by a single nominal scale variable. This scale is nominal because the different digital formats cannot be rank ordered in any way.

Which types of assistive technologies blind and visually impaired persons use both for desktop computers as well as on mobile and wearable devices is measured by a dichotomous variable for each assistive technology type item with the two categories yes or no because a person concerned can use multiple assistive technology types simultaneously at a time. What product for each type of assistive technology they are using is measured by a single nominal scale variable because the different products cannot be rank ordered in any way.

Which input modalities blind and visually impaired persons use for the reading and navigation within structured documents both for desktop computers as well as on mobile and wearable devices is measured by a dichotomous variable with the two categories yes or no for each input modality item because a person concerned can use multiple input modalities simultaneously at a time.

What logical structure elements blind and visually impaired persons use for the navigation within structured documents and how important each of these logical structure elements are for them is measured by a Likert scale because it consists of multiple items. For each logical structure element item an ordinal scale variable with the 4 categories unimportant, less Important, important and very important is employed. This scale is ordinal because these categories can be rank ordered according to their importance where unimportant is less important then very important.

Which accessibility problems blind and visually impaired persons have as to the reading of structured documents and how often each of these accessibility problems occur to them is measured by a Likert scale because it consists of multiple items. For each accessibility problem item an ordinal scale variable with the 4 categories never, seldom, often and very often is employed. This scale is ordinal because these categories can be rank ordered according to their frequency where never is less frequent then very often.

What accessibility solutions blind and visually impaired persons have to solve the problem at the moment if a person concerned is unable to read a specific structured document or parts of it because the document is inaccessible is measured by a single nominal scale variable. This scale is nominal because the different accessibility solutions cannot be rank ordered in any way.

Which hardware devices blind and visually impaired persons are using for the reading of structured documents at the moment and which hardware devices they are planning to purchase in the future within the next 2 to 3 years is measured by a dichotomous variable with the two categories yes or no for each hardware device item because a person concerned can use multiple hardware devices simultaneously at a time.

To which of the 6 visual impairment category according to the WHO International Classification of Diseases ICD (WHO, 1994) the research subjects belong is measured by an ordinal scale variable with the categories mild, moderate, severe, blind, blind with light perception and fully blind without light perception. This scale is ordinal because the different visual impairment categories can be rank ordered according to their severity where mild is less severe than fully blind. If the persons concerned also have a visual field restriction in addition to the restricted visus is measured by a dichotomous scale variable with the two categories yes or no and if they are visually impaired by birth is measured by a dichotomous scale variable with the two categories yes or no too.

The computer affinity the research subjects come with is measured by an ordinal scale variable with the 4 categories none, little, good and very good. This scale is ordinal because these categories can be rank ordered according to their goodness where very good is better than none.

In which occupational situations the research subjects are is measured by a nominal scale variable where each occupational situation is reflected by one category. This scale is nominal because the different occupational situations cannot be rank ordered in any way.

What educational background the research subjects have is measured by an ordinal scale variable where each category reflects the highest education level. This scale is ordinal because the different highest education levels can be rank ordered according to their hight where university doctorate is higher than elementary school.

If the research subjects are interested in being involved in further research by participating in a pilot study or by testing of a prototype of a novel concept for an assistive technology in support of blind and visually impaired persons is measured by a dichotomous scale variable with the two categories yes or no.

3.2.4 Sampling

The sample is a segment of the population that has been selected for investigation. It is a subset of the population. The sampling methods of selection employed in this research are based on a probability as well as on a non-probability approach.

The population of interest for this research from which the sample has to be selected are all blind and visually impaired persons of the world. The World Blind Union holds records and provides on their web-site a directory with all organisations in support of blind and visually impaired persons in each country of the world (WBU, 2015). This directory has been used as the sampling frame for the selection of the participating countries and organisations.

For the survey used to create the user interface Germany, Switzerland and Austria had been selected. For the evaluation the selection has been extended to the United Kingdom, and the United States. These countries had been selected randomly using simple random sampling as probability sampling method. This probability sampling method is likely to generate a representative sample which reflects the population accurately so that it is a microcosm of the population from which it has been taken and sampling error is kept to a minimum. Therefore external validity is high and findings can be generalised to the population because each unit in the population has the same chance of being selected.

In each selected country all main organisations as well as their sub-organisations for each county had been contacted and provided with an invitation letter with the request to send out this invitation to all of their members by distributing the invitation letter across their mailing lists, news letters, forums, magazines and websites. In that invitation, the goal of the research was introduced to them by a short description, followed by a link to the website of the survey questionnaire. This method of selecting the research subjects is called convenience sampling. In addition in this invitation letter the members are asked to forward this invitation to as many of their friends as possible. This sampling method is called snowball sampling.

The limitation of these non-probability sampling methods is that the resulting sample may not be a representative sample of the population and hence it may not be possible to generalize the findings to the population although the results are statistically significant because the research subjects had not been selected randomly as with a probability sampling method. This implies that some units in the population are more likely to be selected than others. For example in case of the convenience sample as mentioned above they all are members of the same organisations or in case of the snowball sample they all know each other. Therefore external validity is typically low.

However this non-probability sampling methods are the only way possible to sample blind and visually impaired subjects for this research because there is no directory of all blind and visually impaired persons in the world and the organisations are not allowed to give out the records of their members for data protection reasons. The organisations can only send out an invitation letter to all of their members and if a member is interested in participating in this research he or she can get in contact by responding to this invitation letter. Therefore it is not possible to actively select the research subjects employing probability sampling methods like simple random sampling as used for the selection of the participating countries. Instead the sample is passively formed of all respondents who replied to the invitation letter.

A source of non-sampling error is non-response. This occurs when some members of the sample refuse to cooperate. For example if an organisation does not forward the invitation letter to its members or a member does not respond to the invitation letter, cannot be contacted or cannot supply the required data for example, because of mental incapacity. Invitations sent out over the Internet are prone to poor response rates. Some sampling errors derive from the fact that there is a large dark number of blind and visually impaired persons who are not members of any organisation and therefore they are unreachable for this research.

The exact sample size cannot be determined in advance because the sample is formed passively of all respondents who replied to the invitation letter and it is not possible to actively select the research subjects as described above. It is that absolute size of a sample that is important not its relative size. This means that increasing the size of a sample increases the precision of a sample. It means that the 95 per cent confidence interval narrows. However a large sample cannot guarantee precision. Increasing the size of a sample increases the likely precision of a sample. This means that, as sampling size increases, sampling error decreases.

3.2.5 Data Collection

A self-completion questionnaire has been employed as the research instrument for collecting the data over the Internet. Before administering this self-completion questionnaire to the sample it has been tested in a pilot study using structured interviews conducted in person on a face-to-face basis as well as over telephone.

Self-Completion Questionnaire

With a self-completion questionnaire, respondents answer questions by completing the questionnaire by themselves. The self-completion questionnaire used for this research as appended in Appendix A consists of the following main parts:

First the introductory statement is located at the beginning and gives a detailed description of the research, containing all relevant information, a prospective research subject requires to decide, if he or she would like to participate in this survey or not. This includes: Goal of the research, procedure, data protection, contact details of the researcher as well as of the supervisory team, informed consent and ethical approval..

After that introduction for each of the 15 concepts and research question as described in Chapter 3.2.1 both pro-coded fixed-choice closed questions as well as open questions are presented. First for each concept or research question a set of so called closed, closed ended, pre-coded, or fixed-choice answers are presented. With a closed question they are presented with a set of pre-coded fixed-choice closed answers alternatives from which they have to choose an appropriate answer. Afterwards an open question with a free text field for additional answers which are not contained within the multiple fixed-choices scale is presented which are post-coded afterwards. Within each group of questions, general questions precede specific ones.

All closed questions are in the forced-choice format. When asking a question that allows the respondent to select more than one answer and where there is the potential that more than one answer are equivalent, a conventional fixed-choice format is used too because there is compelling evidence that the forced-choice format is superior to asking the respondent to tick all answers that apply. Smyth, Dillman, Christian & Stern (2006) have shown the the forced-choice format results in more options being selected. As a result, Dillman, Phelps, Tortora, Swift, Kohrell, Berck & Messer (2009) advocate the use of the forced-choice format for this kind of question situation.

The advantages of closed questions are that they are easy for interviewers and respondents to complete and that it is easy to process the answers. In addition they enhance the comparability of answers. Closed questions may clarify the meaning of a question for respondents. Sometimes, respondents may not be clear about what a question is getting at, and the availability of answers may help to clarify the situation for them. The disadvantages of closed questions are that there is a loss of spontaneity in the respondents answers. There is always the possibility that they might come up with interesting replies that are not covered by the fixed answers that are provided. It can be difficult to make fixed-choice answers manually exclusive so that they do not overlap. It is also difficult to make fixed-choice answers exhaustive. A category of other may be desirable to provide a wide range of answers. There may be variation among respondents in the interpretation of fixed-choice answers. Closed questions may be irritating to respondents when they are not able to find a category that they feel applies to them.

The advantage of open questions on the other hand are that respondents can answer however they wish. Respondents can answer in their own terms. They are not forced to answer in the same terms as those foisted on them by the response choices. They allow unusual responses to be derived. Replies that may not have been contemplated and that would therefore not form the basis for fixed-choice answers are possible. The questions do not suggest certain kinds of answer to respondents. They are useful for exploring new areas or ones in which there is limited knowledge. They are useful for generating fixed-choice format answers. The disadvantages of open questions are that they are timeconsuming for interviewers to administer, the answers have to be post-coded and that they require greater effort from respondents.

At the end of the self-completion questionnaire optional contact details like E-Mail address, postal address and phone number, which may be used for further questions or for the participation in further research, may be provided. The research subjects are acknowledged for taking part in this research and their valuable assistance with which they made a big contribution to the research in support of blind and visually impaired persons and they are informed how they can obtain a copy of the published results which will be made available for download over the Internet or will be sent out to the participants by E-Mail.

This self-completion questionnaire has been applied to the sample over the Internet. The procedure, each research subject had to undertake for completing this self-completion questionnaire consists of the following steps: In that invitation letter sent out to the members of the sample by the different selected organisations in support of blind and visually impaired persons as described in the previous Chapter 3.2.4, the goal of the research was introduced to them by a short description, followed by a link to the website of the self-completion questionnaire. The deadline for answering the questionnaire has been set to the next 30 days. If the prospective research subject is interested in participating in this research, he or she enters the website of the self-completion questionnaire by following the link provided within the invitation letter. The participant answers the questionnaire and submits the form by using the button provided at the end of the questionnaire. The answering of the questionnaire took about 30 minutes. If the participant did not complete all mandatory forced-choice questions, an appropriate message is displayed to him, giving the opportunity to complete the missing information and to submitting the questionnaire again.

Structured Interviews Pilot Study

It is always desirable to conduct a pilot study before administering a self-completion questionnaire to the sample. In fact, the desirability of piloting such instruments is not solely to do with trying to ensure that survey questions operate well. Piloting also has a role in ensuring that the research instrument as a whole functions well.

In this research the pilot study has been used for the following reasons: Open questions were asked in the pilot to generate the fixed-choice answers for the self-completion questionnaire. If anyone or virtually everyone who answers a question replies in the same way, the resulting data are unlikely to be of interest because they do not form a variable. In interview surveys, it may be possible to identify questions that make respondents feel uncomfortable and to detect any tendency for respondents' interest to be lost at certain junctions. Questions that seem not to be understood are more likely to be realized in an interview than in self-completion questionnaire context or questions that are often not answered should become apparent. Pilot studies allow the determination of the adequacy of instructions to respondents completing a self-completion questionnaire. It may be possible to consider how well the questions flow and if it is necessary to move some of them around to improve the feature.

For piloting the self-completion questionnaire described above and testing the questions a pilot study in the form of structured interviews had been carried out on 20 blind and visually impaired research subjects which are not part of the sample employed in the full study but are comparable to members of the population from which the sample for the full study had been taken. These structured interviews had been conducted on a face-to-face as well as over telephone. The self-completion questionnaire above had been used as the interview schedule research instrument for the structured interviews.

A structured interview entails the administration of an interview schedule by an interviewer. The aim is for all interviewees to be given exactly the same context of questioning. This means that each respondent receives exactly the same interview stimulus as any other. The goal of this style of interviewing is to ensure that interviewees' replies can be aggregated, and this can be achieved reliably only if those replies are in response to identical cues. Interviewers are supposed to read out questions exactly and in the same order as they are printed on the schedule. Questions are usually very specific and very often offer the interviewee a fixed range of answers. This type of question is often called closed, closed ended, pre-coded, or fixed-choice.

The advantages of the self-completion questionnaire over the structured interview are: It is cheaper to administer and quicker to administer, absence of interviewer effect. Various studies have demonstrated that characteristics of interviewers and respondents may affect the answers the people give. Obviously since there is no interviewer present when a self-completion questionnaire is being completed, interviewer effects are eliminated. No Interviewer variability. Self-completion questionnaires do not suffer from the problem of interviewers asking questions in a different order or different way. Convenience for respondents: Self-completion questionnaires are more convenient for respondents, because they can complete a questionnaire when they want and at the speed that they want to go.

The disadvantages of the self-completion questionnaire over the structured interview are: There is no interviewer present to prompt and to help respondents if they are having difficulty answering a question. There is no opportunity to probe respondents to elaborate an answer. Probing can be very important when open-ended questions are being asked. It is not possible to ask many questions that are not salient to respondents because respondents are more likely than in interviews to become tired of answering questions that are not very salient to them, and that they are likely to perceive as boring. In addition to ensuring that not too many questions that are not salient to respondents are asked, as previously suggested, it is also important to avoid asking more than a very small number of open questions because respondents frequently do not want to write a lot. Respondents are able to read the whole questionnaire before answering the first question. When this occurs, none of the questions asked is truly independent of the others. It also means that it is not sure that the questions have been answered in the correct order. It is unknown who and whether the right person has answered the questionnaire. Additional data cannot be collected. Because of the possibility of respondent fatigue, it is not possible to ask a lot of questions and large questionnaires are rarely feasible. They may even result in a greater tendency for questionnaires not to be answered in the first phase, since they can be off-putting. Respondents whose literacy is limited will not be able to answer the questionnaire. Partially answered questionnaires are more likely because of a lack of prompting or supervision, than in interviews. It is also easier for respondents actively to decide not to answer a question when on their own than when being asked by an interviewer. For example, questions that appear boring or irrelevant to the respondent may be especially likely to be skipped. If questions are not answered, this creates the problem of missing data for the variables that are created. One of the most damaging limitations is that surveys by self-completion questionnaire typically result in lower response rates than comparable interview based studies.

There are several advantages of telephone interviews over personal interviews: On a like-for-like basis they are far cheaper and quicker to administer. The telephone interview is easier to supervise than the personal interviews. Telephone interviewing has a further advantage, which is to do with evidence that suggests that, in personal interviews, respondents' replies are sometimes affected by characteristics of the interviewer and in deed by his mere presence. Implying that the interviewees may reply in ways they feel will be deemed desirable by interviewers. The remoteness of the interviewer in telephone interviewing removes this potential source of bias to a significant extent because the interviewers personal characteristics cannot be seen, and the fact the he is not physically present may offset the likelihood of respondents' answers being affected by the interviewer.

Telephone interviewing suffers from certain limitations compared to the personal interview: People who do not own or are not contactable by telephone obviously cannot be interviewed. Respondents with hearing impairment are likely to find telephone interviewing much more difficult. The length of a telephone interview is unlikely to be sustainable beyond 20-25 minutes. There is a general belief that telephone interviews achieve slightly lower response-rates than personal interviews. There is some evidence to suggest that telephone interviews fare less well for the asking of questions about sensitive issues. Telephone interviewers cannot engage in observation. This means that they are not in a position to respond to signs of puzzlement or unease on the face of respondents when they are asked a questions. There is some evidence to suggest that the quality of data derived from telephone interviews is inferior to that of comparable face-to-face interviews.

3.2.6 Data Processing

Coding

Coding is a key stage in quantitative research. Answers to open questions are essentially in an unstructured form. In order to quantify and analyse such answers they have to be coded. This unstructured material must be categorised. The answers of the research subjects must be examined and grouped into different categories.

Coding an open question usually entails reading respondents replies and formulating distinct themes in their replies. A coding frame then needs to be designed that identifies the types of answer associated with each question. A coding schedule needs may also be necessary to keep a record of rules to be followed in the identification of certain kinds of answer in terms of a theme.

According to Bryman & Cramer (2011), when coding, three basic principles need to be observed: The categories that are generated must not overlap. The list of categories must be complete and therefore cover all possibilities. It it is not, some material will not be capable of being coded. This sometimes includes a category of “other”.There should be clear rules about how codes should be applied. This ensures that those who are conducting the coding are consistent over time in how they assign material to categories.

The term coding frame is often employed to describe the lists of codes that should be applied to unstructured data and the rules for their application. In structured observation the term coding manual is often preferred to describe the lists of codes for each item of information and the rules to be employed. When indicators are used that are not true quantities, they need to be coded to be turned into quantities.

3.2.7 Data Analysis

The variables created by collecting and processing the data of the different attitudinal and personal factual concepts and research questions outlined in Chapter 3.2.1 and measured as described in Chapter 3.2.2 had been analysed employing the following data analysis methods:

First each variable itself (univariate) had been analysed. Univariate analysis refers to the analysis of one variable at a time. Likert scale ordinal variables are represented by using diverging stacked bar charts with percentages. Each bar represent the number of research subjects falling in each category. Figure 3.2 shows an example of this analysis:

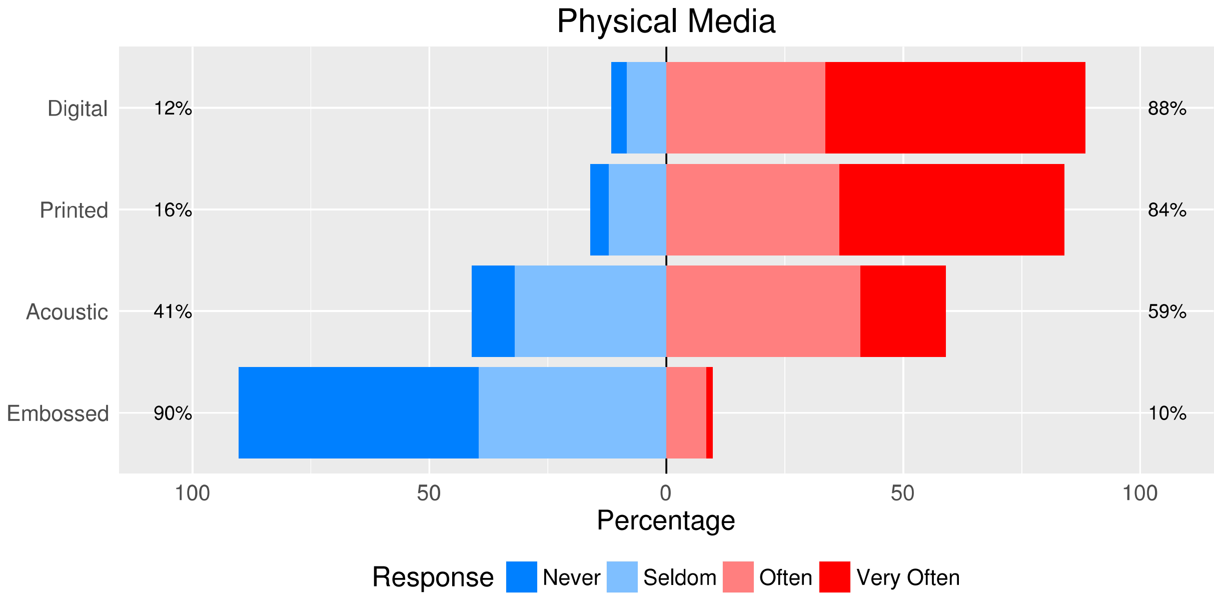


Figure 3.2: Univariate Analysis of Likert Scale Ordinal Variables

If the variable is a regular nominal, ordinal or dichotomous scale variable, rainbow coloured pie charts with frequencies and percentages are used to present that kind of data. This also shows the relative size of the different categories but brings out as well the size of each slice relative to the total sample. The frequency and percentage that each slice represents of the whole sample is also given in this diagram. Figure 3.3 shows an example of this analysis:

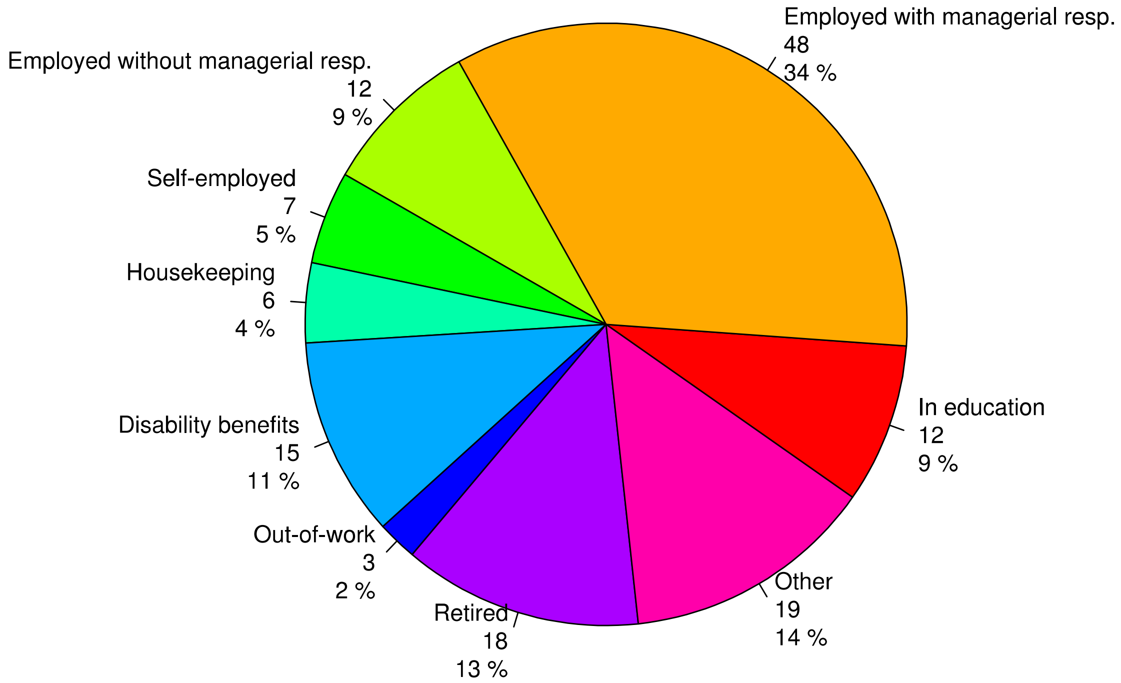


Figure 3.3: Univariate Analysis of regular Nominal or Ordinal Scale Variables

Afterwards each variable has been paired with each other variable in order to analyse the bivariate correlation of the two variables. For each pair of variables Cramer's V (v) strength of the correlation, coefficient of determination (cd), Pearson's Chi-squared (x^2) test of statistical significance, degrees of freedom (df) and probability (p) is calculated.

Bivariate analysis is concerned with the analysis of two variables at a time in order to undercover whether or not the two variables are related. Exploring relationships between variables means searching for evidence that the variation in one variable coincides with variation in another variable. A variety of techniques is available for examining relationships, but their use depends on the nature of the two variables being analysed.

The Cramer’s V (v) coefficient is used for the analysis of the strength of the relationship between two nominal, ordinal or dichotomous variables. It results in a computed statistic that varies between 0 meaning zero or no relationship between the two variables and 1 which represents a perfect relationship. The closer the coefficient is to 1, the stronger the relationship and the closer it is to 0, the weaker is the relationship. A perfect relationship, which would have a Cramer’s V correlation of 1 means that, as one variable increases, the other variable increases or decreases by the same amount and that no other variable is related to either of them. If the correlation is below 1, it means that the first variable is related to at least one other variable as well as to the second variable. The Cramer’s V statistic can take on only a positive value, so that it can get an indication only of the strength of the relationship between two variables and not of the direction.

The coefficient of determination (cd) is a further useful statistic derived by squaring the value of Cramer’s V. It expresses how much of the variation in one variable is due to the other variable. In this research, it is expressed as percentage by multiplying it by 100. The coefficient of determination is a useful adjunct to the interpretation of correlation information.

As we saw in Chapter 3.2.4, there is always the possibility that sampling error, the difference between the population and the sample that has been selected, has occurred, even when probability sampling procedures have been followed. A test of statistical significance allows to estimate how confident we can be that the results deriving from a study based on a randomly selected sample are generalizable to the population from which the sample was drawn. When examining statistical significance in relation to the relationship between two variables, it also tells us about the risk of concluding that there is in fact a relationship in the population when there is no such relationship in the population. If an analysis reveals a statistically significant finding, this does not mean that the finding is intrinsically significant or important. The word significant seems to imply importance. However, statistical significance is solely concerned with the confidence researcher can have in their findings. It does not mean that a statistically significant finding is substantively significant.

The level of statistical significance is the level of risk prepared to take inferring that there is a relationship between two variables in the population from which that sample was taken when in fact no such relationship exists. The maximum level of risk that is conventionally taken in social research is to say that there are up to 5 chances in 100 that we might be falsely concluding that there is a relationship when there is not one in the population from which the sample was taken. This means that, if we draw 100 samples, we are recognising that as many as 5 of them might exhibit a relationship when there is not one in the population. Our sample might be one of that 5, but the risk is fairly small. This significance level is denoted by $p < 0.05$, where p means probability, or by 95 per cent. If we accepted a significance level of $p < 0.1$ or 90 per cent, we would be accepting the possibility that as many as 10 in 100 samples might show a relationship where none exists in the population. In this case there is a greater risk than with $p < 0.05$ that we might have a sample that implies a relationship when there is not one in the population, since the probability of our having such a sample is greater when the risk is 1 in 10 than when the risk is 1 in 20. Therefore, we would have greater confidence when the risk of falsely inferring that there is a relationship between two variables is 1 in 20, as against 1 in 10. But, if a more stringent test is required, perhaps we are worried about the use that might be made of the results, the $p < 0.01$ level might be chosen. This means that we are prepared to accept as the level of risk a probability of only 1 in 100 that the results could have arisen by chance that is, due to sampling error.

The Pearson's chi-square (x^2) test is applied to nominal, ordinal or dichotomous variables. It allows us to establish how confident we can be that there is a relationship between the two variables in the population. The test works by calculating for each pair of categories of the variables an expected frequency or value that is, one that would occur on the basis of chance alone. The chi-squared (x^2) value is produced by calculating the differences between the actual and expected values for each pair of categories of the variables and than summing those differences. The chi-squared value means nothing on its own and can be meaningfully interpreted only in relation to its associated level of statistical significance. Whether or not a chi-square value achieves statistical significance depends not just on its magnitude but also on the number of categories of the two variables being analysed. This latter issue is governed by what is known as the degrees of freedom (df) associated with the two variables. The number of degrees of freedom is governed by multiplying the number of categories minus one of each variable. In other words, the chi-squared value that is arrived at is affected by the number of categories of the two variables, and this is taken into account when deciding whether the chi-square value is statistically significant or not.

Examining the statistical significance of a computed correlation coefficient, which is based on a randomly selected sample, provides information about the likelihood that the coefficient will be found in the population from which the sample was taken. Whether or not a correlation coefficient is statistically significant or not will be affected by two factors: The size of the computed coefficient and the size of the sample. This second factor may appear surprising. Basically, the larger a sample, the more likely it is that a computed correlation coefficient will be found to be statistically significant. Because the question of whether or not a correlation coefficient is statistically significant depends so much on the sample size, it is important to realize that always both the correlation coefficient and the significance level should be examined and one should not be examined at the expense of the other.

In addition the correlation between the two variables is graphically represented. If the significant dependent variable is a Likert scale ordinal variable, grouped diverging stacked bar charts with percentages are used. Figure 3.4 shows an example of this visualisation:

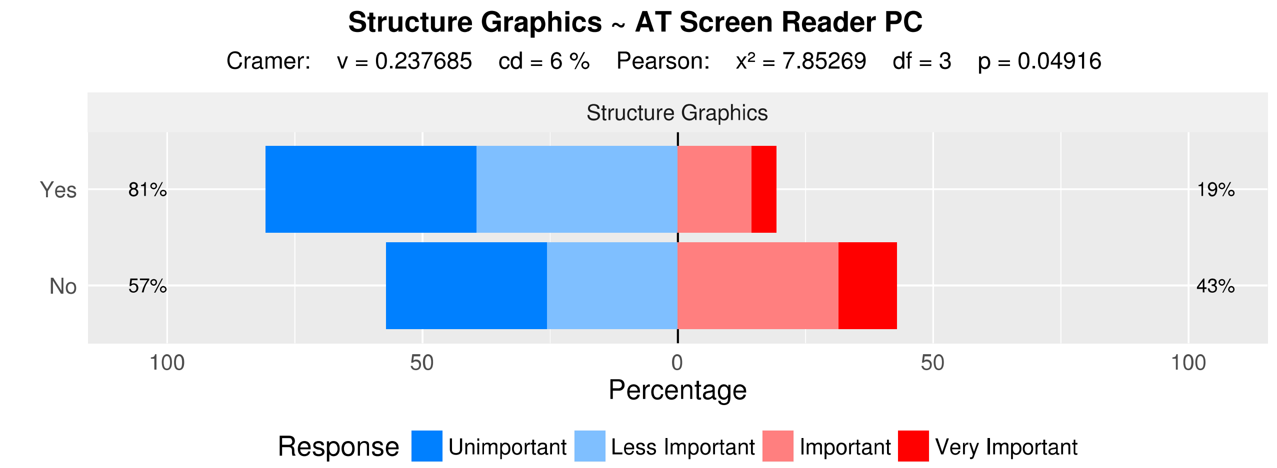


Figure 3.4: Bivariate Analysis of Likert Scale Ordinal Variables

If the significant dependent variable on the other hand is a normal nominal or ordinal scale variable rainbow coloured grouped stacked bar charts with frequencies and percentages are used. Figure 3.5 shows an example of this visualisation:

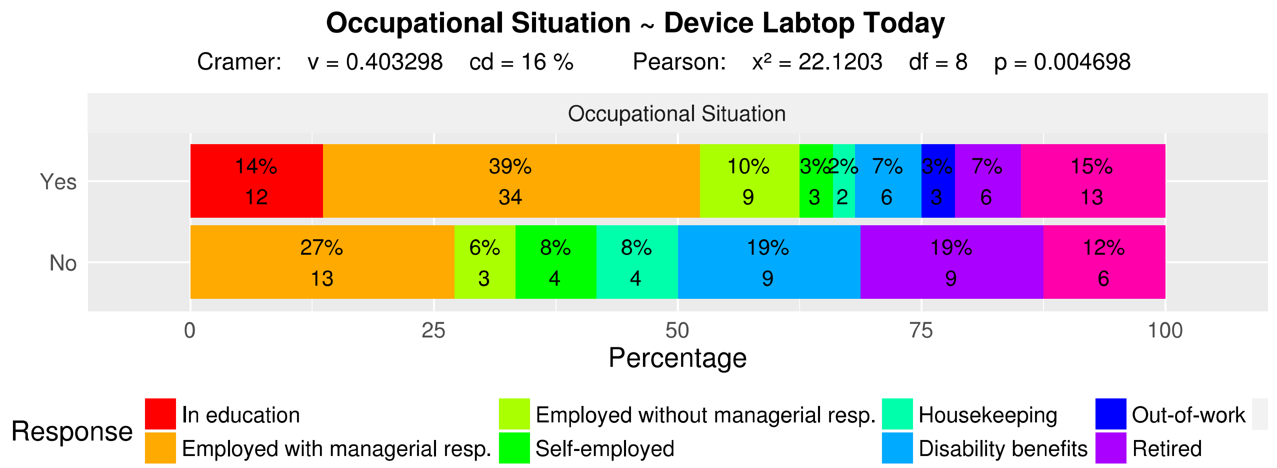


Figure 3.5: Bivariate Analysis of regular Nominal or Ordinal Scale Variables

In this research, only bivariate correlations where the strength of the correlation according to the table of Envans (1991) is at least moderate (v >= 0.4) and with a minimal confidence level of statistical significance of 95% (p <= 0.05) had been accepted.

3.3 Evaluation

For this evaluation of the user interface, a deductive research approach has been used. The purpose of theory is to generate the hypotheses that can be tested and that will thereby allow explanations of laws to be assessed. This is the principle of deductivism. Hence data are collected to test the theories, in this case the user interface design and the user interaction design. Deduction represents an alternative strategy for linking theory and research. Theory and the hypotheses deduced from it come first and drive the process of gathering data. Deductive theory represents the commonest view of the nature of relationship between theory and social research. On the basis of what is known about in a particular domain and of theoretical considerations in relation to the domain, hypotheses are deduced that must than be subjected to empirical scrutiny. Embedded within the hypothesis will be concepts that will need to be translated into researchable entities what are so called operational terms.

3.3.1 Hypotheses

On the basis of the user interface design as described in Chapter 5.2 and the user interaction design as described in Chapter 5.3 as well as of theoretical considerations in relation to the domain, the following hypotheses that must be subjected to empirical scrutiny had been deduced. Embedded within the hypothesis are concepts that had been translated into researchable entities which are so called operational terms.

The main hypothesis is that the proposed user interface design concepts and user interaction design concepts will be effective means for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices. If they are not then there would be no point in using them and alternative methods would have to be found. It may be that they are too complex for users to understand.

In this research, the following 7 user interface design concepts and their hypotheses about the behaviour of blind and visually impaired persons among structured documents in relation to the actions for navigation provided by the different input modalities which are the Multitouch Input Modality as described in Chapter 5.2.5 and the Orientation Motion Input Modality as described in Chapter 5.2.6 had been subjected Table 3.3:

Concept	Hypotheses
Set Structure Cursor **Move Structure Cursor** **Set Text Cursor** **Move Text Cursor** **Unset Text Cursor** **Unset Element Cursor** **Set Select Modifier**	1. **InputModality~Device:** Persons who are using mobile devices like smart-phones and smart-tablets will choose to use the multitouch input modality to perform the navigation action in question because these category of devices typically have a large multi-touch screen embedded and may be too large and heavy to be moved around themselves for using motion. On the other hand, persons who are using wearable devices like smart-watches will choose to use the motion input modality because these category of devices typically are very small and lightweight and can therefore be moved around easily. In addition there may be only a very small multi-touch screen or multi-touch pad or no multi-touch screen or multi-touch pad at all embedded in these devices making the use of multi-touch difficult. 2. **Effectiveness:** The user interface design concepts for performing the navigation action in question by using one of the different input modalities will be effective and enable blind and visually impaired persons to perform the navigation action in question by using one of the different input modalities within a usable amount of time taken and with a usable effort of number of other performed interactions. 3. **Effectiveness~InputModality:** The same effectiveness (time taken and number of interactions performed) to performing the navigation action in question will be achieved across all different input modalities because with crossmodal interaction, the different modalities are used to present the same data. This provides a common representation of the data from all modalities (Gibson, 1966) making them congruent informationally (Marks, 1978). Crossmodal use of the different modalities will allow the characteristics of one sensory modality to be transformed into stimuli for another sensory modality. 4. **Effectiveness~ScreenReader:** Blind and visually impaired persons who are using screen reader will perform better than blind and visually impaired persons who are not using screen reader because neuro-psychological results suggest the assumption that blind persons can use acoustic information more effectively than sighted persons and that areas of the cortex responsible for visual information also can process acoustical information (Schick & Meis, 2002). These results suggest that blind persons can at least use acoustical navigational informations as well as sighted persons (Donker, Klante & Gorny, 2002). 5. **Effectiveness~Device:** The same effectiveness (time

	taken and number of interactions performed) to performing the navigation action in question will be achieved across all the different hardware devices employed because the results of Hoggan & Brewster (2010) show that there are times when audio feedback is more appropriate than tactile feedback and vice versa. For this reason devices which support both, tactile feedback and audio feedback, will cover the widest range of platforms, environments, applications, user preferences, locations and tasks. 6. **Time~Interactions:** The two direct indicators, the time taken and the number of interactions performed, can be used as a set of multiple indicators to measure the effectiveness of the proposed user interface design concepts for performing the navigation action in question by using one of the different input modalities as follows: The more effective the navigation user interface design concepts in question are, the less time will be taken and the less interaction will be performed.

Table 3.3: Navigation User Interface Design Concepts and Behavioural Hypotheses

In addition to the user interface design concepts for navigation actions as outlined above, the following 6 user interface design concepts and their hypotheses about the behaviour of blind and visually impaired persons among structured documents in relation to the reactions for the presentation of the element attributes and text provided by the different output modalities which are the Earcon Output Modality as described in Chapter 5.2.2, the Tacton Output Modality as described in Chapter 5.2.3 and the Speech Output Modality as described in Chapter 5.2.4 had been subjected Table 3.4:

<table>
<tr><th>Concept</th><th>Hypotheses</th></tr>
<tr><td>Element Type</td><td rowspan="6">1. InputModality~Device: Persons who are using mobile devices like smart-phones and smart-tablets will choose to use the multitouch input modality to recognise the element attribute or text in question because these category of devices typically have a large multi-touch screen embedded and may be too large and heavy to be moved around themselves for using motion. On the other hand, persons who are using wearable devices like smart-watches will choose to use the motion input modality because these category of devices typically are very small and lightweight and can therefore be moved around easily. In addition there may be only a very small multi-touch screen or multi-touch pad or no multi-touch screen or multi-touch pad at all embedded in these devices making the use of multi-touch difficult.

2. Effectiveness: The user interface design concepts for presenting the element attribute or text in question to the user by the reactions provided by the different output modalities will be effective and enable blind and visually impaired persons to recognise the element attribute or text in question as presented by the reactions of the different output modalities with a recall rate which will be much better than the number of selects required to recognise the element attribute or text in question by chance only and within a usable amount of time taken and with a usable effort of performed interactions.

3. Effectiveness~InputModality: The same effectiveness (number of selects performed, time taken and number of interactions performed) to recognise the element attribute or text in question will be achieved across all different input modalities because with crossmodal interaction, the different modalities are used to present the same data. This provides a common representation of the data from all modalities (Gibson, 1966) making them congruent informationally (Marks, 1978). Crossmodal use of the different modalities will allow the characteristics of one sensory modality to be transformed into stimuli for another sensory modality.

4. Effectiveness~ScreenReader: Blind and visually impaired persons who are using screen reader will perform better than blind and visually impaired persons who are not using screen reader because neuro-psychological results suggest the assumption that blind persons can use acoustic information more effectively than sighted persons and that areas of the cortex responsible for visual information also can process acoustical information (Schick & Meis, 2002). These</td></tr>
<tr><td>Level</td></tr>
<tr><td>Position</td></tr>
<tr><td>Length</td></tr>
<tr><td>Relationship</td></tr>
<tr><td>Text</td></tr>
</table>

	results suggest that blind persons can at least use acoustical navigational informations as well as sighted persons (Donker, Klante & Gorny, 2002). 5. **Effectiveness~Device:** The same effectiveness (number of selects performed, time taken and number of interactions performed) to recognise the element attribute or text in question will be achieved across all the different hardware devices employed because the results of Hoggan & Brewster (2010) show that there are times when audio feedback is more appropriate than tactile feedback and vice versa. For this reason devices which support both, tactile feedback and audio feedback, will cover the widest range of platforms, environments, applications, user preferences, locations and tasks. 6. **Selects~Time+Interactions:** The three direct indicators, the number of selects performed, time taken and the number of interactions performed, can be used as a set of multiple indicators to measure the effectiveness of the proposed user interface design concepts for presenting the element attribute or text in question by the reactions provided by the different output modalities as follows: The more effective the presentation user interface design concepts in question are, the less selects will be performed, the less time will be taken and the less interactions will be performed.

Table 3.4: Presentation User Interface Design Concepts and Behavioural Hypotheses

In addition to the user interface design concepts for navigation and presentation as outlined above, the following 6 user interaction design concepts and their hypotheses about the behaviour of blind and visually impaired persons among structured documents in relation to the interactions (use cases) for manipulation provided by the user interaction design as described in Chapter 5.3 had been subjected Table 3.5:

Concept	Hypotheses
Activate Selection **Deactivate Selection** **Move Selection** **Remove Selection** **Insert Element** **Insert Text**	1. **InputModality~Device:** Persons who are using mobile devices like smart-phones and smart-tablets will choose to use the multitouch input modality to perform the manipulation interaction in question because these category of devices typically have a large multi-touch screen embedded and may be too large and heavy to be moved around themselves for using motion. On the other hand, persons who are using wearable devices like smart-watches will choose to use the motion input modality because these category of devices typically are very small and lightweight and can therefore be moved around easily. In addition there may be only a very small multi-touch screen or multi-touch pad or no multi-touch screen or multi-touch pad at all embedded in these devices making the use of multi-touch difficult. 2. **Effectiveness:** The user interaction design concept for performing the manipulation interaction in question provided by the user interaction design will be effective and enable blind and visually impaired persons to perform the manipulation interaction in question by using one of the different input modalities within a usable amount of time taken and with a usable effort of number of other performed interactions. 3. **Effectiveness~InputModality:** The same effectiveness (time taken and number of interactions performed) to performing the manipulation interaction in question will be achieved across all different input modalities because with crossmodal interaction, the different modalities are used to present the same data. This provides a common representation of the data from all modalities (Gibson, 1966) making them congruent informationally (Marks, 1978). Crossmodal use of the different modalities will allow the characteristics of one sensory modality to be transformed into stimuli for another sensory modality. 4. **Effectiveness~ScreenReader:** Blind and visually impaired persons who are using screen reader will perform better than blind and visually impaired persons who are not using screen reader because neuro-psychological results suggest the assumption that blind persons can use acoustic information more effectively than sighted persons and that areas of the cortex responsible for visual information also can process acoustical information (Schick & Meis, 2002). These results suggest that blind persons can at least use acoustical navigational informations as well as sighted persons (Donker, Klante & Gorny, 2002). 5. **Effectiveness~Device:** The same effectiveness (time

	taken and number of interactions performed) to performing the manipulation interaction in question will be achieved across all the different hardware devices employed because the results of Hoggan & Brewster (2010) show that there are times when audio feedback is more appropriate than tactile feedback and vice versa. For this reason devices which support both, tactile feedback and audio feedback, will cover the widest range of platforms, environments, applications, user preferences, locations and tasks. 6. **Time~Interactions:** The two direct indicators, the time taken and the number of interactions performed, can be used as a set of multiple indicators to measure the effectiveness of the proposed user interaction design concept for performing the manipulation interaction in question by using one of the different input modalities as follows: The more effective the manipulation user interaction design concept in question is, the less time will be taken and the less interactions will be performed.

Table 3.5: Manipulation User Interaction Design Concepts and Behavioural Hypotheses

In addition to the user interface design concepts for navigation and presentation as well as the user interaction design concepts for manipulation and their behavioural hypotheses as outlined above, the following 5 personal factual concepts and research questions about personal information of the research subjects and the hardware device used by them to participate in this research, had been subjected Table 3.6:

Concept	Research Questions
Region	• In which region of the world is the device of the research subject located ?
Language	• Which language did the device of the research subject request ?
Operating System	• Which operating system is the device of the research subject running ?
Device	• What type or hardware device does the research subject use in order to participate in this research ?
Screen Reader	• Does the research subject have a screen reader such as Apple [VoiceOver] or Google [Talkback] enabled on the hardware device used ?

Table 3.6: Personal Factual Concepts and Research Questions

3.3.2 Measures of Concepts

Measures can be taken to refer to concepts that can be relatively unambiguously counted or in other words, measures are quantities. To tap concepts that are less directly quantifiable indicators that will stand for the concept are needed. An indicator, than, is something that is devised or already exists and that is employed as though it were a measure of a concept. It is viewed as an indirect measure of a concept. In order to provide a measure of a concept often referred to as an operational definition, it is necessary to have an indicator or multiple indicators that will stand for the concept.

In this research, the 13 user interface design concepts for navigation and presentation and the 6 user interaction design concepts for manipulation and their hypotheses about the behaviour of blind and visually impaired persons among structured documents as well as the 5 personal factual concepts and research questions about personal information of the research subjects and the hardware device used by them to participate in this research as described in Chapter 3.3.1 are measured as follows:

Effectiveness

If and how well blind and visually impaired persons are able to perform a navigation action provided by the different input modalities, recognise an element attribute or text presented by the reactions of the different output modalities or to perform a manipulation interaction provided by the user interaction design in question is not directly measurable. Because this is not directly measurable indicators are required to measure the effectiveness of these concepts. They are too complex to rely on only one indicator because a single indicator may incorrectly classify many individuals or may capture only a portion of the underlying concepts. Therefore a set of multiple indicators are used. The following 3 direct indicators (number of selects performed, time taken and number of interactions performed) were used to measure the effectiveness of these user interface design concepts and user interaction design concepts as follows:

Number of Selects

How many selects the research subjects have to perform in order recognise an element attribute or text in question is directly measured by an interval ratio variable where each category represents a distinct number of selects. This is a direct indicator. The less selects a respondent require, the more effective the underlying user interface or user interaction design concept in question works.

Time

How much time in seconds [s] it takes for the research subjects to perform a navigation action, recognise an element attribute or text or to perform a manipulation interaction is directly measured by an interval ratio scale variable where each category represents one second. This is a direct indicator. The less time a respondent require, the more effective the underlying user interface or user interaction design concept in question works.

Number of Interactions

How many interactions the research subjects have to perform in order to performing a navigation action, recognise an element attribute or text or to perform a manipulation interaction is directly measured by an interval ratio variable where each category represents a distinct number of interactions. This is a direct indicator. The less interactions a respondent require, the more effective the underlying user interface or user interaction design concept in question works.

Input Modality

Which input modality the research subjects have chosen in order to perform a navigation action, recognise an element attribute or text or to perform a manipulation interaction in question is directly measured by a nominal scale variable with the two categories Multitouch or Motion. This scale is nominal because these different input modalities cannot be rank ordered in any way.

Region

In which region of the world the hardware devices of the research subjects are located is directly measured by a nominal scale variable where each category represents one of the selected countries as described in Chapter 3.2.4 which are Germany, Switzerland, Austria, the United Kingdom and the United States as well as a category of other which represents countries which were not directly invited. These other countries are grouped to a single category of other. This scale is nominal because these different countries cannot be rank ordered in any way.

Language

Which language the hardware devices of the research subjects requested is directly measured by a nominal scale variable with the categories English and German as well as a category of other because the Doky Structured Observation App is provided in English and German language. If another language is requested, English is served because it is the default language. These other languages are grouped to a single category of other. This scale is nominal because these different languages cannot be rank ordered in any way.

Operating System

Which operating system the hardware devices of the research subjects are running is directly measured by a nominal scale variable with the categories iOS and Android because the Doky Structured Observation App is provided for the Apple [iOS] as well as for the Google [Android] operating system. This scale is nominal because these different operating systems cannot be rank ordered in any way.

Hardware Device

Which type of hardware device the research subjects are employing in order to participate in this research is directly measured by an ordinal scale variable with the categories Smart-Watch, Smart-Phone and Smart-Tablet. This scale is ordinal because these different types of hardware device can be rank ordered according to their size where Smart-Watch is the smallest and Smart-Tablet is the largest device.

Screen Reader

If the research subjects have a screen reader such as Apple [VoiceOver] or Google [TalkBack] enabled on their hardware devices used to participate in this research is directly measured by a dichotomous scale variable with the two categories Yes or No.

3.3.3 Data Collection

For collecting the data for this evaluation, an automated structured observation of the blind and visually impaired research subjects forming the sample as described in Chapter 3.3 performing for each of the user interface design or user interaction design concepts as described in Chapter3.3.1 an exercise among a highly structured example document using the Doky Structured Observation App as described in Appendix C on their own mobile or wearable device over the Internet has been employed.

Structured Observation

When overt behaviour is the focus of analysis and issues of meaning are less salient, structured observation is almost certainly more accurate and effective than getting people to report an their behaviour through questionnaires. When survey techniques such as the structured interview or the self-completion questionnaire as described in Chapter 3.2.5 are employed in connection with the study of respondents behaviour, certain characteristic difficulties are encountered: People may vary in their interpretations of key terms in a question (problem of meaning). When answering the question, respondents may inadvertently omit key terms in the question (problem of omission). They may misremember aspects of the occurrence of certain forms of behaviour (problem of memory). They may exhibit a tendency towards replying in ways that are meant to be consistent with their perceptions of the desirability of certain kinds of answer (social desirability effect). Some questions may appear threatening and result in a failure to provide a honest reply. How people say they are likely to behave and how they actually behave may be inconsistent and can create a gap between stated and actual behaviour. An obvious solution to the problems identified is to observe people’s behaviour directly rather than to rely on research instruments like questionnaires to elicit such information.

Structured observation is a technique in which explicitly formulated rules for the observation and recording of behaviour are employed. The rules inform observers about what they should look for and how they should record behaviour. Each person who is part of the research is observed for a predetermined period of time using the same rules. These rules are articulated in what is usually referred to as an observation schedule, which bears many similarities to a structured interview schedule with closed questions.

Observation Schedule

The aim of the observation schedule is to ensure that each participant’s behaviour is systematically recorded so that it is possible to aggregate the behaviour of all those in the sample in respect of each type of behaviour being recorded. The rules that constitute the observation schedule should be as specific as possible in order to direct observers to exactly what aspects of behaviour they are supposed to be looking for. The resulting data resemble questionnaire data considerably, in that the procedure generates information on different aspects of behaviour that can be treated as variables.

The observation schedule used for this research as appended in Appendix B consists of the following parts. Figure 3.6 gives an overview.

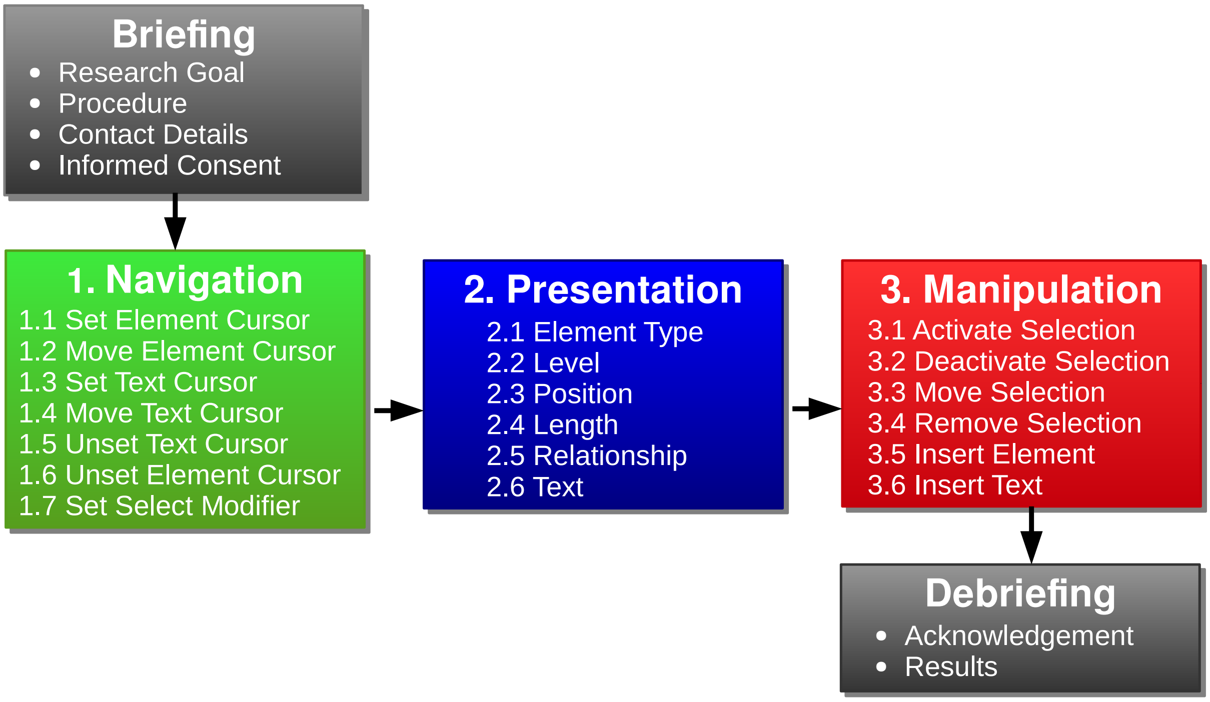


Figure 3.6: Structured Observation Schedule Overview

First the briefing part as the introductory statement is located at the beginning and gives a detailed description of the research, containing all relevant information, a prospective research subject requires to decide, if the person would like to participate in this structured observation or not. This includes: Goal of the research, procedure, data protection, contact details of the researcher as well as of the supervisory team, informed consent and ethical approval.

After this introduction, there are 19 exercises organised in 3 parts. For each user interface design or user interaction design concept as described in Chapter 3.3.1 an exercise in which the research subject has to apply the concept in question by performing a task among a highly structured example document.

The first part consists of the exercises testing the navigation user interface design concepts as described in Table 3.3. In this first 7 exercises the research subject will learn how to navigate within a document. There are two different ways how this can be done: They can either use the Multitouch Input Modality as described in Chapter 5.2.5 by using their fingers on the screen. Or he or she can use the Motion Input Modality as described in Chapter 5.2.6 by moving the device itself. In the instructions always both ways are explained to the research subjects. Afterwards it is up to them to decide which way to use to solve an exercise.

The second part comprises the exercises focussing on the presentation user interface design concepts as described in Table 3.4. In this next 6 exercises the research subject will learn how the document is presented to them by the different output modalities. There are three different ways in which this is done: They hear tones presented by the Earcon Output Modality as described in Chapter 5.2.2, they can feel vibrations presented by the Tacton Output Modality as described in Chapter 5.2.3 and they can hear speech presented by the Speech Output Modality as described in Chapter 5.2.4. In the instructions always all ways are explained to research subjects. Afterwards it is up to them to decide which ways to use to solve an exercise.

In the third and last part exercises for the manipulation user interaction design concepts (use cases) are performed as described in Table 3.5. In this last 6 exercises the research subjects will learn how to activate, move, remove and insert structure elements and text in a document provided by the user interaction design as described in Chapter 5.3.

After they have completed all exercises they are debriefed. In the debriefing part, they are acknowledged for taking part in this research and their valuable assistance in making a big contribution to the research in support of blind and visually impaired persons. In addition, they are informed how they can obtain a copy of the published results, which will be made available for download over the Internet or will be sent out to the participants by E-Mail.

Each exercise consists of the following parts: First, the instructions inform the research subjects on how the user interface design concept or user interaction design concept in question works, what the participant can do, what action he or she have to take and how the reactions of the system are. Afterwards the task tells the research subject what he or she exactly has to do among the example document. In case the research subject has successfully completed the task, congratulations inform the research subject the he or she has successfully completed the task. Afterwards the continuation tells the research subject what he or she has to do to continue with the next part or exercise. Figure 3.7 shows an example of an exercise:

Instructions:
When you are at an element containing text, you can go into this text:
Move the finger 1 centimeter to the left or to the right.
Or rotate the device vertically anticlockwise 10 degrees.
Task:
Please go into an arbitrary text.
Congratulation:
Congratulations, you have successfully gone into a text.
Continuation:
Please tap with one finger anywhere on the screen to continue with the next exercise.

Figure 3.7: Exercise consisting of Introduction, Task, Congratulation and Continuation

Doky Structured Observation App

For applying the observations schedule as described above to the sample, the Doky Structured Observation App as described in Appendix C in detail has been developed which automates the process of structured observation and enables the research subjects to which the research subject can use on their own mobile or wearable device over remotely over the Internet. The Doky Structured Observation App has been implemented for Apple iOS [iOS] as well as for Google Android [Android]. It can be downloaded from the apple AppStore as well as from the GooglePlay store for free.

After the respondent has downloaded, installed and started the Doky Structured Observation App, the following procedure begins: For each exercise of the observation schedule as described above, the instructions and than the task are read out to the research subject by using speech synthesis. Afterwards the user interface with the example document is enabled and the research subject can start performing the task among the example document. When the research subject successfully completed the goal of the task, the user interface with the example document is disabled and the congratulations text is read out to the research subject using speech synthesis. In addition an applause sound is played. After the congratulation the continuation text is read out to the research subject. While the research subject is solving a task the screen is black because the whole exercise is on a non-visual basis. After the whole observation schedule is completed and after the debriefing the app is terminated.

The research subject can listen to the instructions again at any time by double-tapping with one finger anywhere on the screen. In this case the process of performing the task among the example document is interrupted and the user interface example document disabled ant the instructions as well as the task are read to the research subject again using speech syntheses. Afterwards the user interface with the example document is enabled again and the process of performing the task among the example document continues. In addition the participant can withdraw at any time by pressing the home button. If the participant has withdrawn at a time, the Doky Structured Observation App will be terminated immediately and all collected data will be deleted. It is possible to interrupt the conduction of the structured observation. In this case the current status is saved. It is possible to continue at the same position at a later point in time by starting the Doky Structured Observation App again.

During the whole structured observation process the events of the structured observation are recorded and live streamed over the Internet to a server as described in detail in Appendix C using a secured connection where it is possible to follow the ongoing structured observations live in real-time and the events are also stored in password protected directory for later data processing and data analysis as described in the following chapters.

Before the observation schedule and the Doky Structured Observation App had been applied to the sample remotely over the Internet as described above, it had been tested in a pilot study which had been carried out on a face-to-face basis on 10 blind and visually impaired research subjects which are not part of the sample employed in the full study but are comparable to members of the population from which the sample for the full study had been taken.

The advantages of this automated structured observation over a personal structured observation carried out by an observer in person on a face-to-face basis are: They are cheaper to administer and quicker to administer. Observer effects are eliminated. Since there is no observer present when the structured observation is being carried out, characteristics of the observer cannot affect the behaviour of the research subjects. Automated structured observation does not suffer from the problem of observer variability of observers asking the respondents to performing exercises in a different order or in a different way. They are more convenient for the research subjects, because they can conduct the structured observation where they want, when they want and at the speed that they want to go using their own mobile or wearable device.

The disadvantages of this automated structured observation over a personal structured observation carried out by an observer in personal face-to-face are: There is no observer present to prompt and to help respondents if they are having difficulty solving an exercise. There is no opportunity to probe respondents to elaborate a behaviour. Respondents are more likely to become tired of performing exercises that are not very salient to them, and that they are likely to perceive as boring. It is unknown who and whether the right person has carried out the structured observation. Because of the possibility of respondent fatigue, large observation schedules are rarely feasible. They may even result in a greater tendency for the structured observation not to be carried out in the first phase, since they can be off-putting. There is greater risk of missing data. Partially solved exercises are more likely because of a lack of prompting or supervision. It is also easier for respondents actively to decide not to solve an exercise when on their own than when being asked by an observer. For example, exercises that appear boring or irrelevant to the respondent may be especially likely to be skipped. If exercises are not solved, this creates the problem of missing data for the variables that are created. One of the most damaging limitations is that automated structured observations typically result in lower response rates than comparable personal structured observation based studies.

3.3.4 Data Processing

For data processing a Java application named Doky Processor had been developed which reads in the text files with the recorded events as described in Chapter 3.3.3 and Appendix C returned by the Doky Structured Observation App over the Internet and calculates the variables created by the measures of concept for each case.

It generates a single comma separated (CSV) file containing the variables for each case on one line. This table of the cases and the variables for each case is than used as the basis for data analysis as described in the next Chapter 3.3.5 and read into R.

3.3.5 Data Analysis

The variables created by collecting and processing the data of the different behavioural and personal factual concepts and research questions outlined in Chapter 3.3.1 and measured as described in Chapter 3.3.2 had been analysed employing the following data analysis methods:

For each interval ratio variable like time, number of interactions or number of selects a boxplot with values covering the entire variable range gives an overview. This form of display provides an indication on both central tendency, the median, and dispersion of the range. It also indicates whether there are any outliers. The box represents the middle 50 percent of cases. The upper line of the box indicates the greatest case within the 50 per cent and the lower line of the box represents the least within the 50 per cent. The line going across the box indicates the median. The line going upwards from the box goes up to the case greater than any other, other than the outliers. The line going downwards from the box goes down to the case lower than any other. Boxplots are useful because they display both central tendency and dispersion. Afterwards a histogram with kernel density line, frequencies and percentages describes the detailed distribution of the interesting range of the variable only without the outliers. As with the bar charts, the bars represent the relative size of each of the bands. Figure 3.8 shows an example of this analysis:

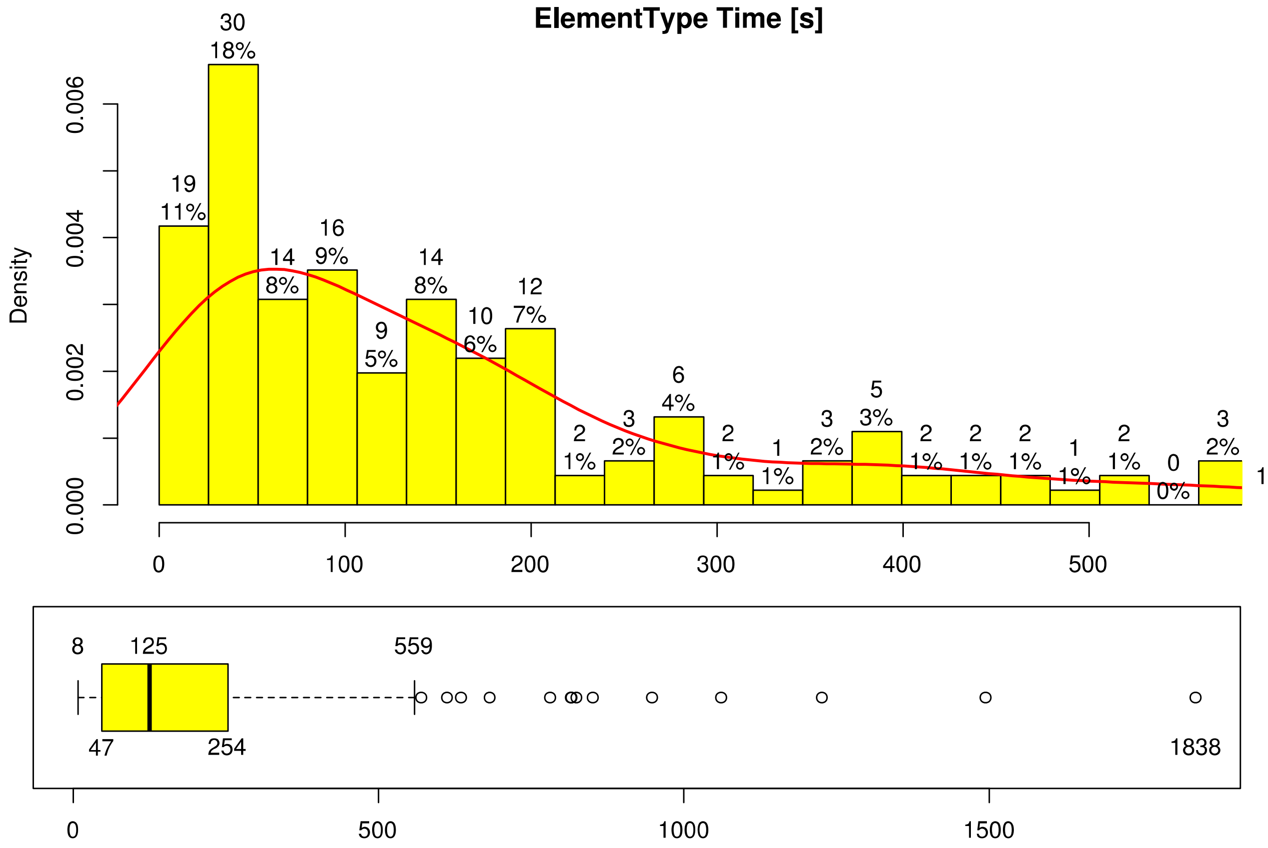


Figure 3.8: Univariate Analysis of Interval Ratio Variables

For each bivariate correlation of an interval ratio variable with a nominal, ordinal or dichotomous scale variable like the input modality used, device used or the use of screen reader eta test of association (eta), coefficient of determination (cd), f-statistic test of statistical significance, t-statistic (t), degrees of freedom (df) and probability (p) is calculated.

To examine the relationship between an interval ratio variable and a nominal, ordinal or dichotomous scale variable, and if the latter can be relatively unambiguously identified as the significant independent variable, a potentially fruitful approach is to compare the means of the interval ratio variable for each category of the nominal, ordinal or dichotomous scale variable. This procedure is accompanied by a test of association between the two variables called eta. This statistic expresses the strength of the correlation of the two variables and, like Cramer's V, will always be positive.

Eta-squared or the coefficient of determination (cd) expresses the amount of variation in the interval ratio variable that is due to the nominal, ordinal or dichotomous scale variable in per cent. Eta is a very flexible method for exploring the relationship between two variables, because it can be employed when one variable is nominal, ordinal or dichotomous and the other interval ratio. Also, it does not make the assumption that the relationship between the variables is linear.

A test of statistical significance can also be applied to the comparison of means that was carried out as described above. This procedure entails treating the total amount of variation in the significant dependent variable as made up of two types: Variation within the categories that make up the significant independent variable, and variation between them. The latter is often called the explained variance and the former the error variance. A test of statistical significance for the comparison of means entails relating the two types of variance to form what is known as the F-statistic. This statistic expresses the amount of explained variance in relation to the amount of error variance. The F-statistic value, as with the Chi-squared value, means nothing on its own and can be meaningfully interpreted only in relation to its associated level of statistical significance. Whether or not an F-statistic value achieves statistical significance depends not just on its magnitude but also on the number of categories of the two variables being analysed. This latter issue is governed by the degrees of freedom (df) associated with the two variables. The number of degrees of freedom is governed by multiplying the number of categories minus one of each variable. In other words, the F-statistic value that is arrived at is affected by the number of categories of the two variables, and this is taken into account when deciding whether the F-statistic value is statistically significant or not.

In addition this bivariate correlation is graphically represented by grouped boxplots with values. Figure 3.9 shows an example of this analysis:

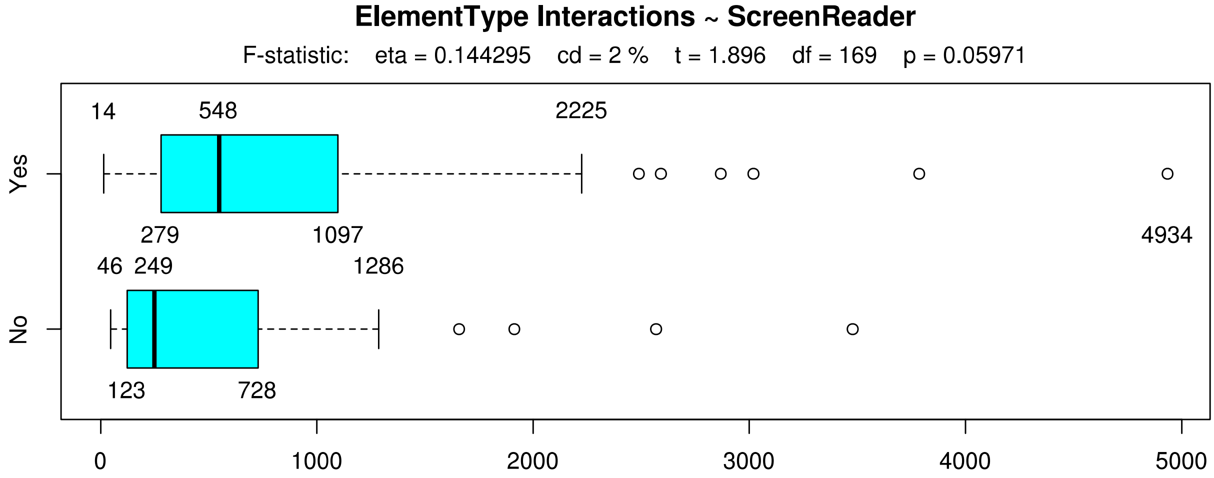


Figure 3.9: Bivariate Analysis of an Interval Ratio and a Nominal or Ordinal Variable

For each correlation between two interval ratio variables Pearson's product moment correlation coefficient (cor) strength of correlation, coefficient of determination (cd), t-statistic (t), degrees of freedom (df) and statistical significance (p) is calculated.

The Pearson's product moment correlation coefficient (cor) is a method for examining relationships between two interval ratio variables. The coefficient will almost certainly lie between 0 meaning zero or no relationship between the two variables and 1 which represents a perfect relationship. This indicates the strength of the relationship. The closer the coefficient is to 1, the stronger the relationship and the closer it is to 0, the weaker is the relationship. The coefficient will be either positive or negative. This indicates the direction of the relationship. A perfect relationship, which would have a Pearson’s correlation of 1 means that, as one variable increases, the other variable increases or decreases by the same amount and that no other variable is related to either of them. If the correlation is below 1, it means that the first variable is related to at least one other variable as well as to second variable.

This type of bivariate correlation is graphically represented using scatter plots with linear regression. A 3D scatter plot with colouring, vertical lines and regression plane shows the multivariate correlation between the three interval ratio variables time, number of interactions and number of selects. Multivariate analysis entails the simultaneous analysis of three or more variables. In order for a relationship between two variables to be established, not only must there be evidence that there is a relationship but the relationship must be shown to be non-spurious. A spurious relationship exists when there appears to be a relationship between two variables, but the relationship is not real. It is being produced because each variable is itself related to a third variable which is known as a confounding or intervening variable. An intervening variable allows to answer questions about the bivariate relationship between variables. It suggests that the relationship between the two variables is not a direct one, since the impart is viewed as occurring via a third variable. Hence the relationship between the two variables is a moderated relationship because it is moderated by this third variable. Figure 3.10 shows an example of this analysis:

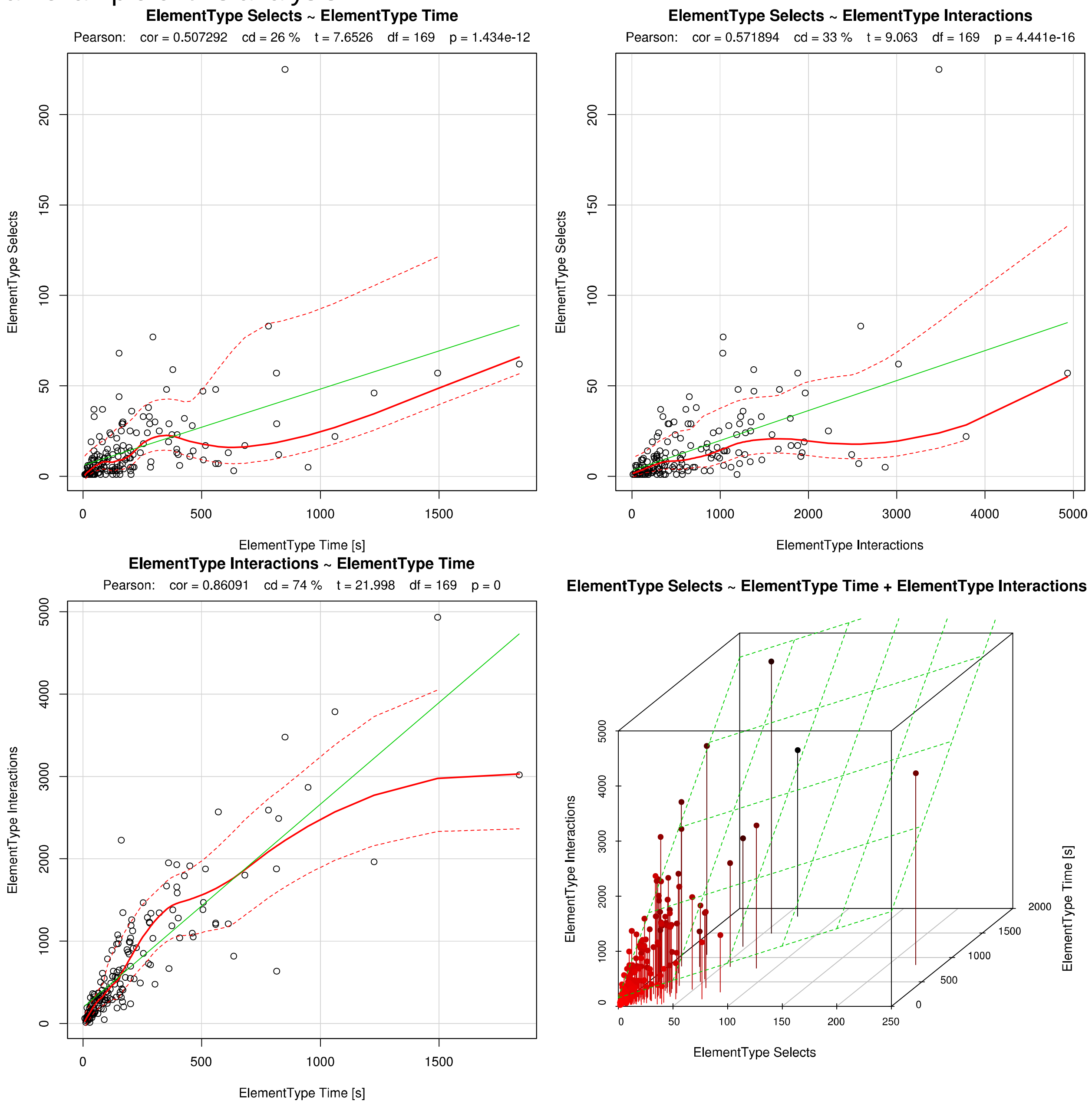


Figure 3.10: Multivariate Analysis between Interval Ratio Variables using Scatter Plots

Nominal, ordinal or dichotomous scale variables are presented using rainbow coloured pie charts with frequencies and percentages as described in Chapter 3.2.7. This also shows the relative size of the different categories but brings out as well the size of each slice relative to the total sample. The frequency and percentage that each slice represents of the whole sample is also given in this diagram. Figure 3.11 shows an example of this visualisation.

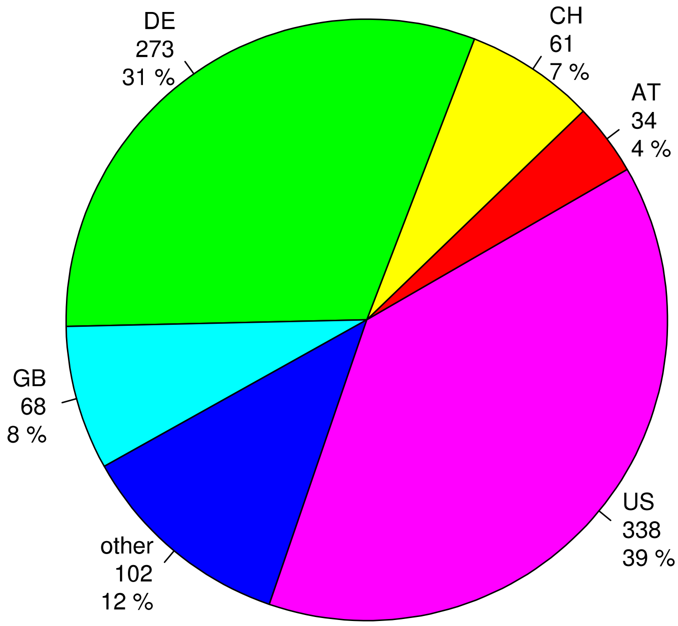


Figure 3.11: Univariate Analysis of Nominal or Ordinal Scale Variables

And for each bivariate correlation between two nominal, ordinal or dichotomous scale variables Cramer's V (v), coefficient of determination (cd), Pearson's Chi-squared (x^2), degrees of freedom (df) and statistical significance (p) is calculated as described in Chapter 3.2.7. In addition this correlation is graphically represented using rainbow coloured stacked bar charts with frequencies and percentages. Figure 3.12 shows an example of this analysis:

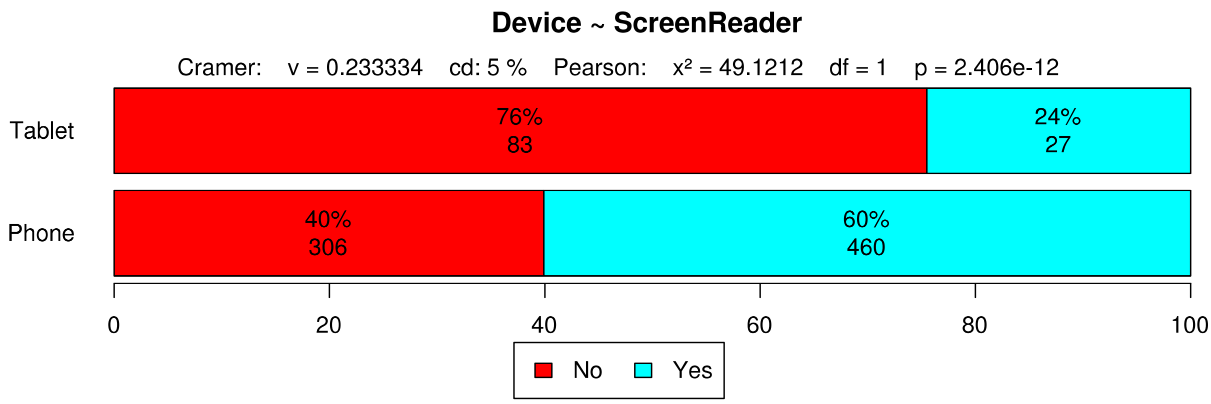


Figure 3.12: Bivariate Analysis of Nominal or Ordinal Scale Variables

3.4 Conclusions

A research methodology for developing the DOKY user interface by employing a natural science epistemological model (positivism), an objective ontology (objectivism), a quantitative research strategy and an iterative research approach has been proposed.

The user interface has been created using an inductive research approach where general research questions stand at the beginning as the hypotheses and the theory, in this case the user interface design and the user interaction design, is the outcome of the research. A cross-sectional survey research design had been employed because more than one case had been examined at a single point in time using quantitative or quantifiable data with more than one variable in order to detect variation between cases and patterns of association between variables. The concepts and their research questions about the attitudes of blind and visually impaired persons in relation to the reading of structured documents as well as the personal factual concepts and research questions and questions about their knowledge in this area are measured using single item nominal, ordinal or dichotomous scale variables as well as multiple-item Likert scale variables. The participating countries and organisations had been selected randomly using simple random sampling as probability sampling method and the research subjects had been selected using convenience sampling and snowball sampling as non-probability sampling methods. A self-completion questionnaire has been employed as the research instrument for collecting the data over the Internet. Before administering this self-completion questionnaire to the sample it has been tested in a pilot study using structured interviews conducted in person on a face-to-face basis as well as over telephone. During the data processing, the answers to open questions had been post-coded. The variables created by collecting and processing the data had been analysed employing diverging stacked bar charts with percentages, pie charts with frequencies and percentages, Cramer's V test of association, coefficient of determination, Pearson's Chi-squared test of statistical significance, degrees of freedom, probability, grouped diverging stacked bar charts with percentages and grouped stacked bar charts with frequencies and percentages.

For the evaluation of the user interface, a deductive research approach has been followed where the purpose of the theory is to generate the hypotheses that can be tested and that will thereby allow explanations of laws to be assessed. The user interface design concepts for navigation and presentation and the user interaction design concepts for manipulation and their research questions as well as the personal factual concepts and research questions about personal information of the research subjects and the hardware device used had been measured using multiple direct indicators with interval ratio scale variables as well as single indicators with nominal, ordinal and dichotomous scale variables. For collecting the data for this evaluation, an automated structured observation of the blind and visually impaired research subjects forming the sample performing for each of the user interface design or user interaction design concepts an exercise among a highly structured example document using the Doky Structured Observation App on their own mobile or wearable device over the Internet has been employed. For data processing a Java application named Doky Processor had been developed which reads in the text files with the recorded events returned by the Doky Structured Observation App over the Internet and generates the variables created by the measures of concepts for each case. The variables created by collecting and processing the data of the different behavioural and personal factual concepts and research questions had been analysed using boxplots with values, histograms with kernel density line, frequencies and percentages, eta test of association, coefficient of determination, f-statistic test of statistical significance, t-statistic, degrees of freedom, probability, grouped boxplots with values, Pearson's product moment correlation coefficient, coefficient of determination, t-statistic, degrees of freedom, statistical significance, scatter plots with linear regression and 3D scatter plots with vertical lines and regression plane.

The last step involved a movement that is in the opposite direction from deduction to induction, as the implications of the findings from the theory that prompted the whole exercise had been inferred. These findings are fed back into the stock of theory and the research findings associated with the domain of non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices.

4 Survey: Structured Documents from the Blind and Visually Impaired Person's Point of View

4.1 Introduction

In order to develop a novel concept for a multi-modal user interface for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices as described in the next Chapter 5, the following survey has been conducted among a significant number of blind and visually impaired persons to find out how they handle structured documents at the moment and what is of importance as to the reading of and the navigation within structured documents from the blind and visually impaired person’s point of view. This survey has been conducted in 2010.

The persons concerned were asked how satisfied they are with their current general situation as to the reading of structured documents, on which different physical media structured documents are provided to them and on which preferred physical medium they would wish to receive their structured documents, in which different digital content formats structured documents are provided to them and in which preferred digital content format they would wish to receive their structured documents, what different types and products of assistive technologies, both for desktop computers as well as on mobile and wearable devices, they are using for the reading of structured documents, which different input modalities they are using for the reading of and the navigation within structured documents, both for desktop computers as well as on mobile and wearable devices, what logical structure elements are of importance for them as to the reading of and the navigation within structured documents, what accessibility problems they have as to the reading of structured documents, what accessibility solutions they are taking on to solve these accessibility problems at the moment and which hardware devices they are using for the reading of structured documents today as well as in the future.

4.2 Research Subjects

The different organisations in support of blind and visually impaired persons in Austria, Germany and Switzerland, which had been selected and invited to participate in this research for sampling the blind and visually impaired research subjects for this survey as described in detail in Chapter 3.2.4, have a potential number of 25‘000 persons concerned as their members.

Out of this potential number of 25‘000 blind and visually impaired persons, feedback from 205 respondents, forming the sample for this research, had been received in the age of 18 to 80 years as well as with different visual performance, different computer affinity, in various different occupational situations, with different educational backgrounds and different genders. This sample is described in more detail in this chapter.

The target number of responses that was hoped to get as baseline was about 100. The exact sample size can not be determined in advance because the sample is formed passively of all respondents who replied to the invitation letter. A source of non-sampling error is non-response. This occurs when some members of the sample refuse to cooperate. For example if an organisation does not forward the invitation letter to it’s members or a member does not respond to the invitation letter, cannot be contacted or can not supply the required data.

4.2.1 Visual Performance

The blind and visually impaired research subjects come with different visual performance. The persons concerned had been grouped into 6 categories of visual impairment according to the WHO International Classification of Diseases (ICD) definition of blind (WHO, 1994). These categories are briefly described below. Figure 4.1 shows the frequencies and percentages of respondents in each of the different categories of visual impairment:

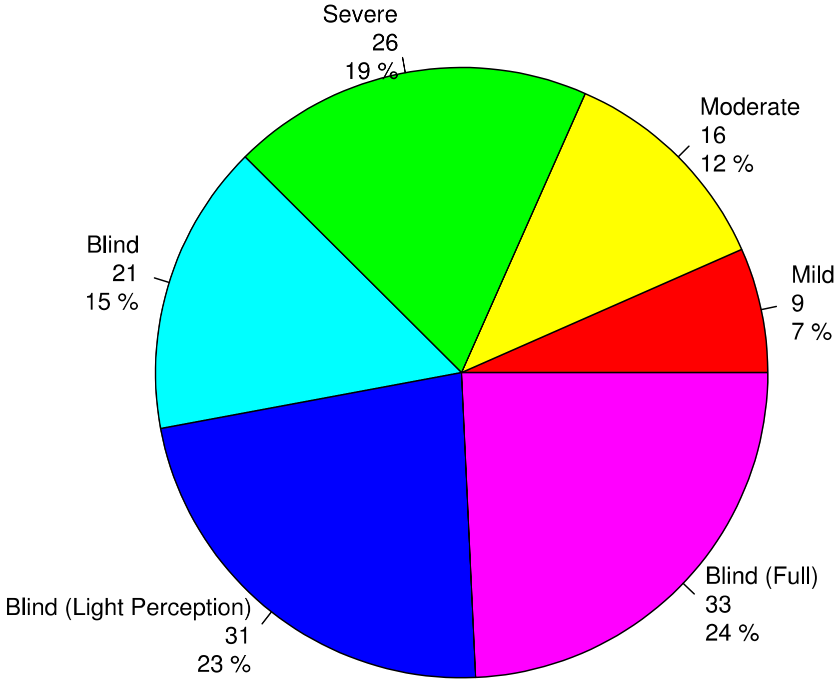


Figure 4.1: Categories of Visual Impairment and Members of Research Subjects

7 per cent of the research subjects suffer from a mild visual impairment with a visus lower than 1.0 and greater than or equal to 0.3. 12 per cent of the respondents suffer from a moderate visual impairment with a visus worse than 0.3 and greater than or equal to 0.1. 19 per cent of the persons concerned suffer from a severe visual impairment with a visus worse than 0.1 and greater than or equal to 0.05. 15 per cent of the participants are legally blind with a visus worse than 0.05. 23 per cent of the research subjects are classified as blind with light perception. This perception of light, namely the ability to distinct between light and dark, is legally defined by a visus of 0.01. Finally, 24 per cent of the persons concerned are classified as fully blind with no light perception. This is legally defined by a visus of 0.0.

In addition to the reduced visual acuity, most research subjects do have a restricted visual field. Figure 4.2 shows the frequencies and percentages of respondents with and without such a visual field restriction:

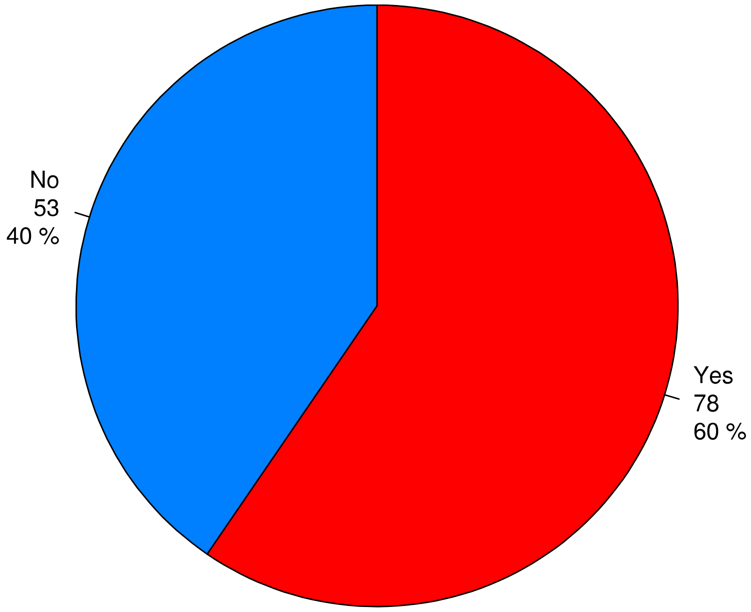


Figure 4.2: Categories of Visual Field Restriction and Members of Research Subjects

60 per cent of the participants have a restricted visual field in addition to the reduced visual acuity and 40 per cent of them do not suffer from a visual field restriction:

Most of the research subjects are visually impaired from birth. This is a very important distinction because a person concerned who has never seen the visual representation of structured documents may handle them different than another person who has been seeing the visual representation of them earlier in childhood. Figure 4.3 shows the frequencies and percentages of respondents who are visually impaired from birth and which had been seeing earlier in childhood:

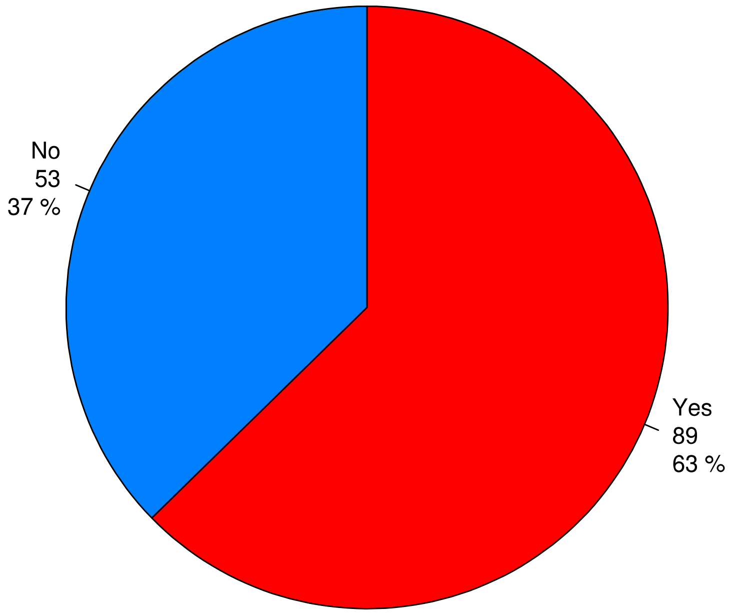


Figure 4.3: Categories of Research Subjects who are Visually Impaired from Birth

63 per cent of the research subjects are visually impaired from birth and only 37 per cent of them had been seeing earlier in childhood.

4.2.2 Computer Affinity

The research subjects come with different computer affinity. The computer affinity a specific person comes with is a very important factor as to the handling of structured documents because it may greatly influence the methods on how the persons concerned navigate within a structured document, which problems they have and how they solve these problems. For this purpose the research subjects had been grouped into the following 4 categories of computer affinity. These categories are briefly described below. Figure 4.4 shows the frequencies and percentages of respondents in each of the different categories of computer affinity:

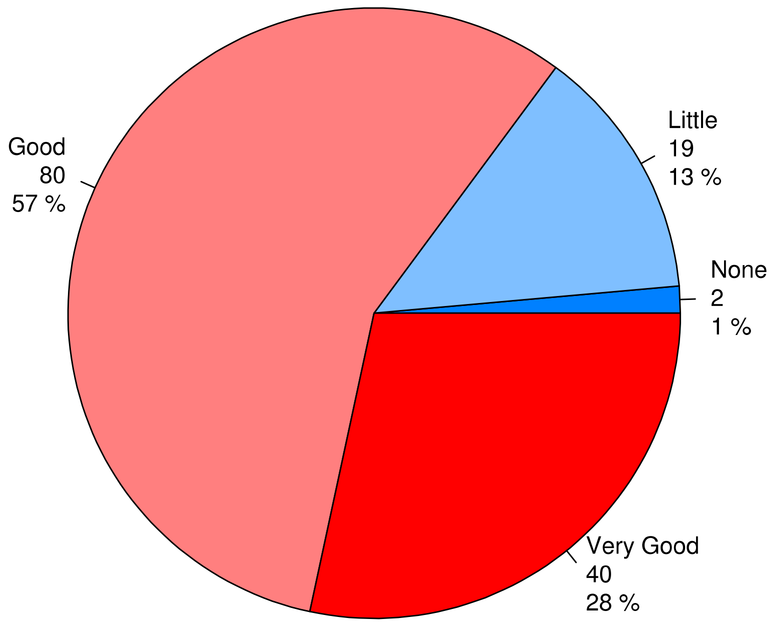


Figure 4.4: Categories of Computer Affinity and Members of Research Subjects

28 per cent of the research subjects classified their computer affinity as very good. Most of the respondents, namely 57 per cent of them, reported to have good computer affinity. 13 per cent of the participants responded to have little computer affinity. Only 1 per cent of the persons concerned claimed not to be computer affine at all. The reason for this very low number of respondents with none computer affinity might be the fact that the data for this research had been collected using as self-completion questionnaire applied to the sample over the Internet as described in detail in Chapter 3.2.5. This method of data collection requires the respondents to have at least little computer affinity and persons with none computer affinity on the other hand, which do not use the Internet at all, were unreachable because they are not able to participate in this research.

4.2.3 Occupational Situation

The research subjects are in different occupational situations. Figure 4.5 shows the different occupational situations reported by the persons concerned as well as the frequencies and percentages of respondents in each occupational situaltion. These occupational situations are briefly described below.

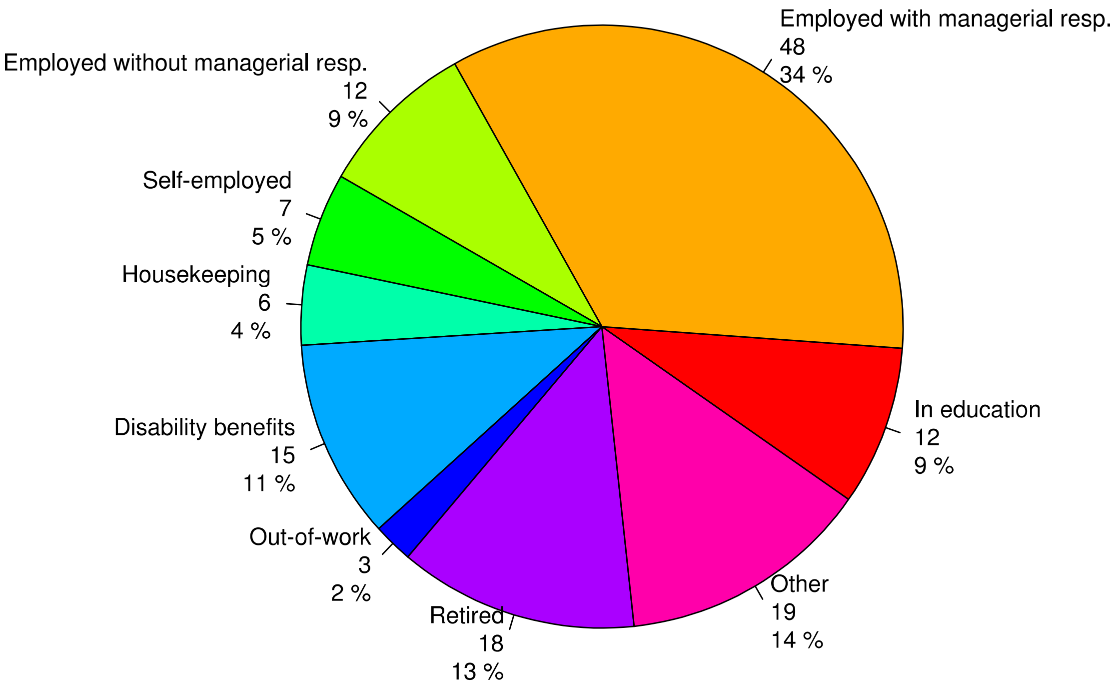


Figure 4.5: Categories of Occupational Situations and Members of Research Subjects

Most of the research subjects, namely 34 per cent of them, are employed with managerial responsibility. 13 per cent of the persons concerned are retired. 11 per cent of the participants receive disability benefits. 9 per cent of the respondents are in education and another 9 per cent of them are employed without managerial responsibility. 5 per cent of the research subjects are working self-employed. 4 per cent of the persons concerned are occupied in housekeeping. Only 2 per cent of the participants are out of work. Finally, 14 per cent of the respondents find themselves in another occupational situation.

4.2.4 Highest Education

The research subjects have different educational backgrounds. Figure 4.6 shows the different educational backgrounds reported by the persons concerned as well as the frequencies and percentages of respondents with each highest education. These educational beckgrounds are briefly described below.

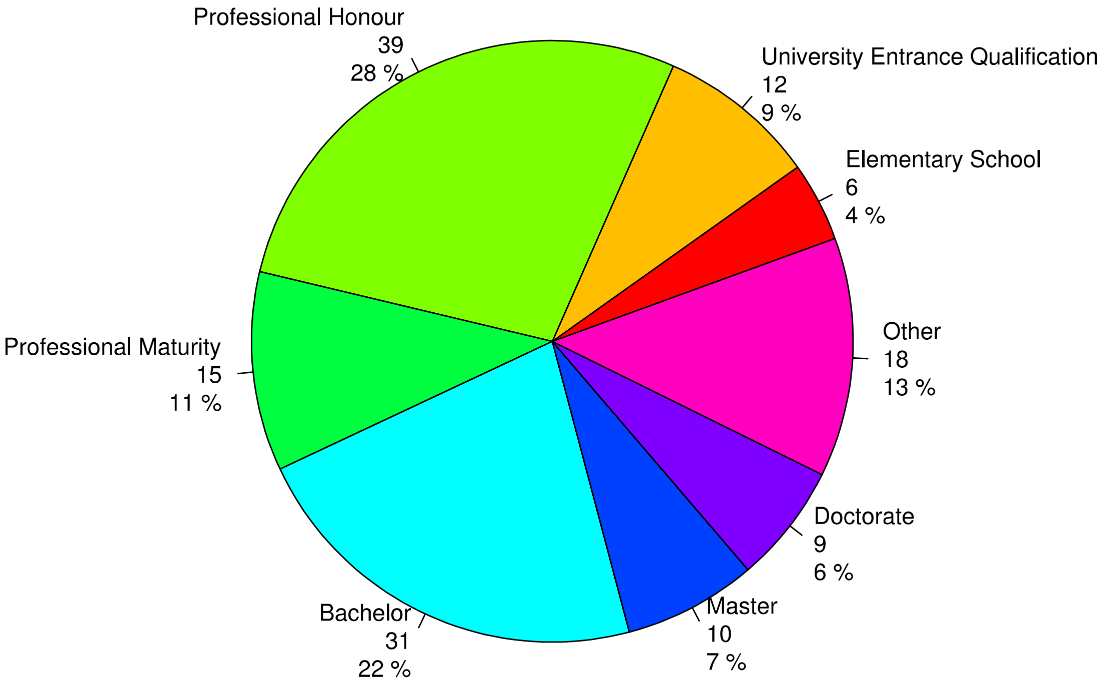


Figure 4.6: Categories of Highest Education and Members of Research Subjects

28 per cent of the research subjects have a professional honor. 22 per cent of the respondents graduated an university diploma or bachelor. 11 per cent of the persons concerned have a professional maturity. 9 per cent of the participants have a general qualification for university entrance. 7 per cent of the research subjects have an university master degree and 6 per cent of them graduated an university doctorate degree. Only 4 per cent of the respondents completed elementary school without higher education. Finally, 13 per cent of the respondents reported to have completed another highest education.

4.2.5 Gender

The research subjects have different genders. Figure 4.7 shows the frequencies and percentages of respondents in each of the two categories of genders:

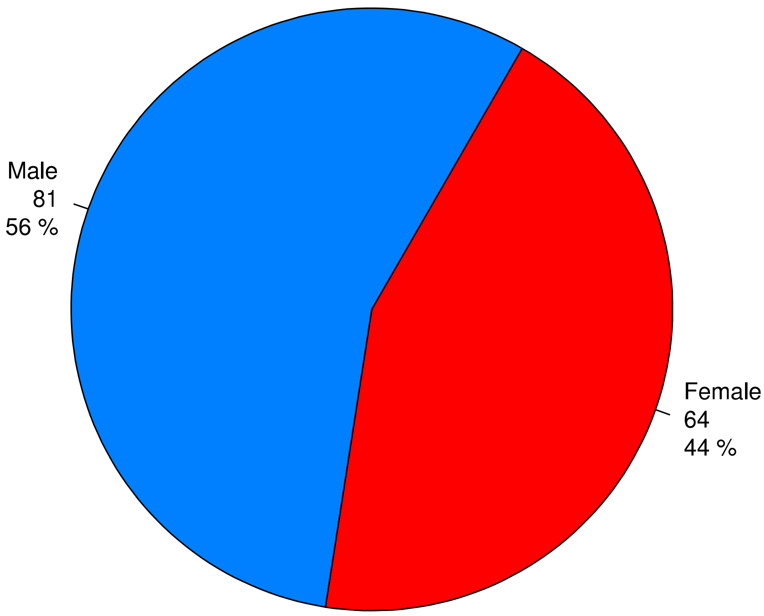


Figure 4.7: Categories of Genders and Members of the Research Subjects

56 per cent of the persons concerned are male and they are accompanied by 44 per cent of female participants.

4.2.6 Further Research

The research subjects are very enthusiastic and interested in being involved in further research. This may involve the participation in a pilot study or the testing of a prototype of a novel concept for an assistive technology in support of blind and visually impaired persons for the reading of structured documents. Figure 4.8 shows the frequencies and percentages of respondents who would like to participate in further research. 61 per cent of the participants would like to be involved in further research and 39 per cent of them are not interested in participating in futher research.

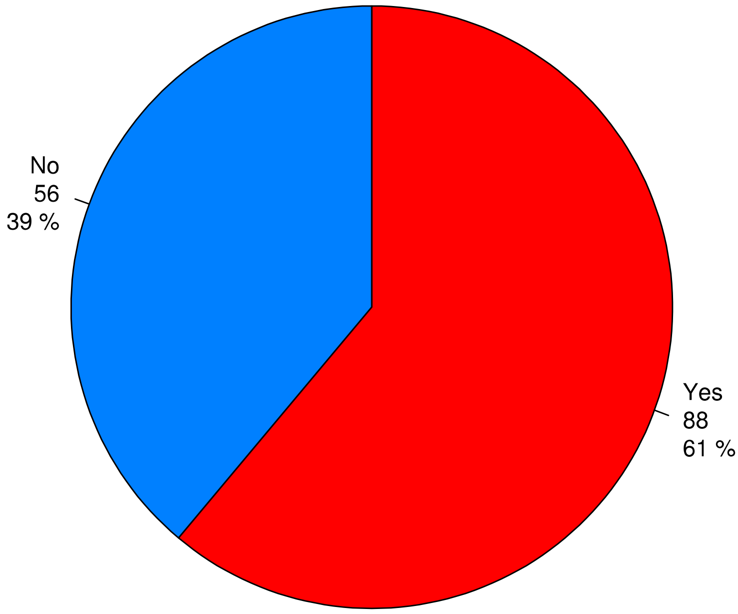


Figure 4.8: Interest of the Research Subjects in the Participation in Further Research

4.3 Results

The feedbacks received from the 205 blind and visually impaired research subjects, who make up the sample for this research as described in detail in the previous Chapter 4.2, lead to the following results for the 10 concepts and their research questions about the attitudes of blind and visually impaired persons in relation to the reading of structured documents which had been subjected in Chapter 3.2.1:

4.3.1 General Satisfaction

At the moment, most blind and visually impaired persons are not very satisfied with their current general situation as to the reading of structured documents. Figure 4.9 shows the different categories of general satisfaction reported by the persons concerned as well as the frequencies and percentages of respondents in each category of general satisfaction. These categories are briefly described below.

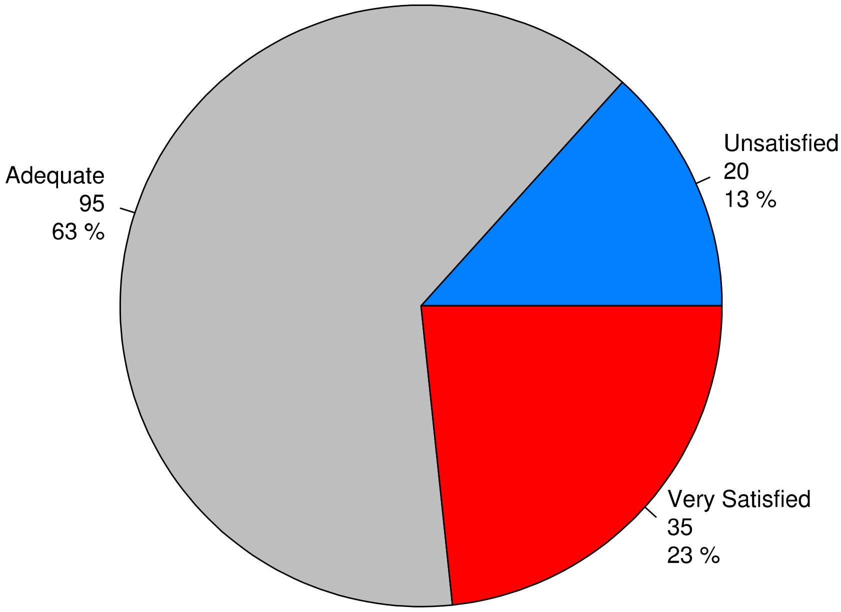


Figure 4.9: General Satisfaction of Research Subjects as to Reading of Documents

Only 13 per cent of the research subjects are very satisfied with their current general situation as to the reading of structured documents. They can read structured documents quick and easy and are able to locate the relevant information fast and they feel informed very well. The general satisfaction of 63 per cent of the respondents is adequate. They can read the absolutely necessary information only and they feel adequate informed. 23 per cent of the persons concerned are unsatisfied with their current general situation. The reading of structured documents is very difficult for them and they always fail in finding the relevant information in time. They feel poor informed.

There is serious need for action required by conducting further research on a novel concept for an assistive technology in support of blind and visually impaired persons for to the reading of structured documents because most of the persons concerned, namely 77 per cent of them, are not very satisfied with their current general situation as to the reading of structured documents.

4.3.2 Physical Media

Structured documents are provided to blind and visually impaired persons at the moment on a mix of different physical media. Figure 4.10 shows the different physical media reported by the persons concerned and the percentages of how often structured documents are provided to the respondents on each physical medium. These different physical media are briefly described below.

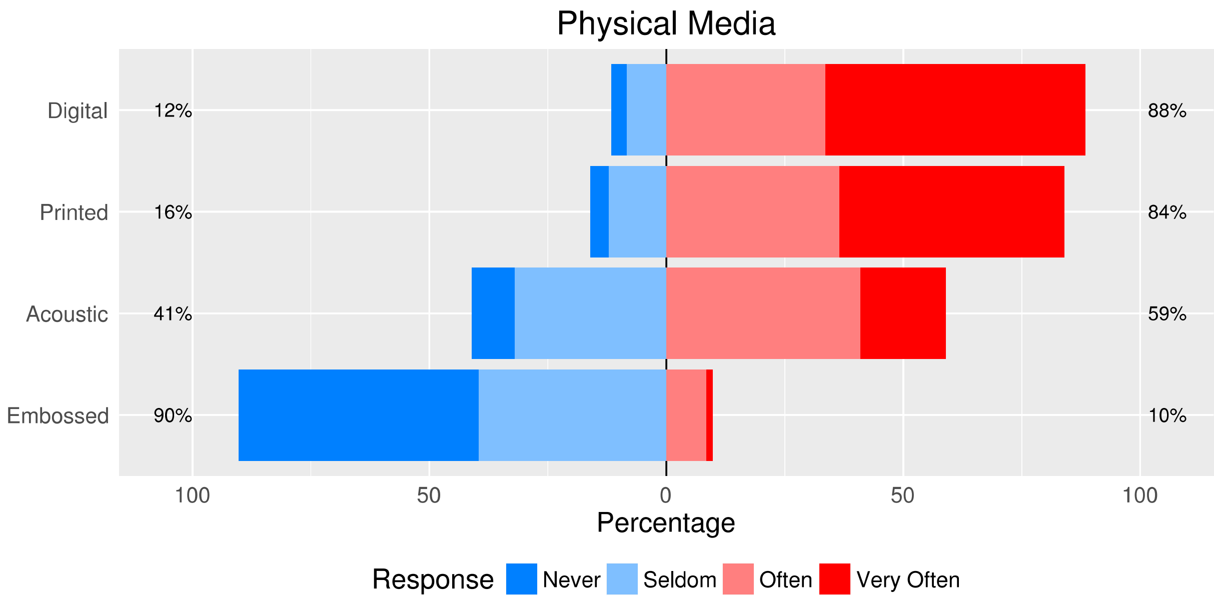


Figure 4.10: Physical Media on which Documents are provided to Research Subjects

88 per cent of blind and visually impaired persons receive structured documents often and very often digitally and only 12 per cent of them receive digital structured documents seldom or never. This fact is due to the emerging digital and web based platforms and services on which structured documents are provided to them, both for desktop computers as well as on mobile and wearable devices.

To 84 per cent of the persons concerned structured documents are provided printed on paper often and very often and only 16 per cent of them receive printed structured documents seldom or never. This is because the printed medium is commonly used both by blind and visually impaired persons as well as by persons without disability.

59 per cent of blind and visually impaired persons receive structured documents often and very often in an acoustic form and 41 per cent of them receive acoustic structured document seldom and never. This acoustic medium is used for example in case of auditory books or podcasts.

If a structured document is embossed on paper, text as well as graphics are represented in tactile form. For text the Braille font is used, where each character is represented by a combination of 6 dots. Graphics are represented tactile by using different dot heights according to the brightness of the visual pixel at an appropriate position. To only 10 per cent of the persons concerned structured documents are provided on this embossed medium and to 90 per cent of them embossed structured documents are provided seldom or never because this is a solution for blind persons only, which is typically not used by persons without disability. Also a combination is possible: A document can be printed and embossed on the same paper at the same time as used by the ViewPlus Tiger Printers [ViewPlus].

In contrast to the current situation, most blind and visually impaired persons would wish to receive their structured documents in digital form. Figure 4.11 shows the different physical media reported by blind and visually impaired persons on which they would wish to receive their structured documents as well as the frequencies and percentages of respondents in each category of physical medium wished:

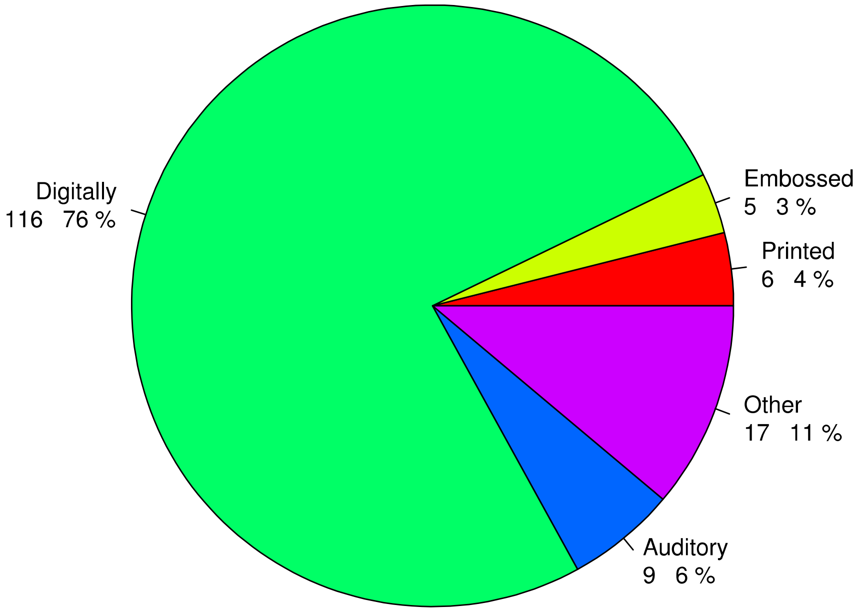


Figure 4.11: Physical Media Research Subjects would wish to receive their Documents

76 per cent of blind and visually impaired persons, would wish to receive their structured documents in digital form. This is because digital structured documents are directly accessible to a variety of different assistive technologies and can easily be adapted to the special needs, for example by changing the font type, size or colour. 6 per cent of the persons concerned would wish to receive structured documents acoustically. 4 per cent of the respondents would wish that structured documents are provided to them printed on paper, 3 per cent embossed on paper and 11 per cent would wish to receive their structured documents on other physical media.

Therefore further research should be investigated in the digital physical medium because in most cases, blind and visually impaired persons would wish to receive their structured documents in digital form because of the emerging digital and web based platforms and services on which structured documents are provided to them and since digital structured documents are directly accessible to a variety of different assistive technologies and can easily be adapted according to the special needs.

4.3.3 Digital Content Formats

At the moment, structured documents are provided to blind and visually impaired persons, as with the physical media described in Chapter 4.3.2 above, in a mix of different digital content formats. Figure 4.12 shows the different digital content formats reported by the persons concerned and the percentages of how often structured documents are provided to the respondents in each digital content format. These different digital content formats are described in detail in Chapter 2.2.3.

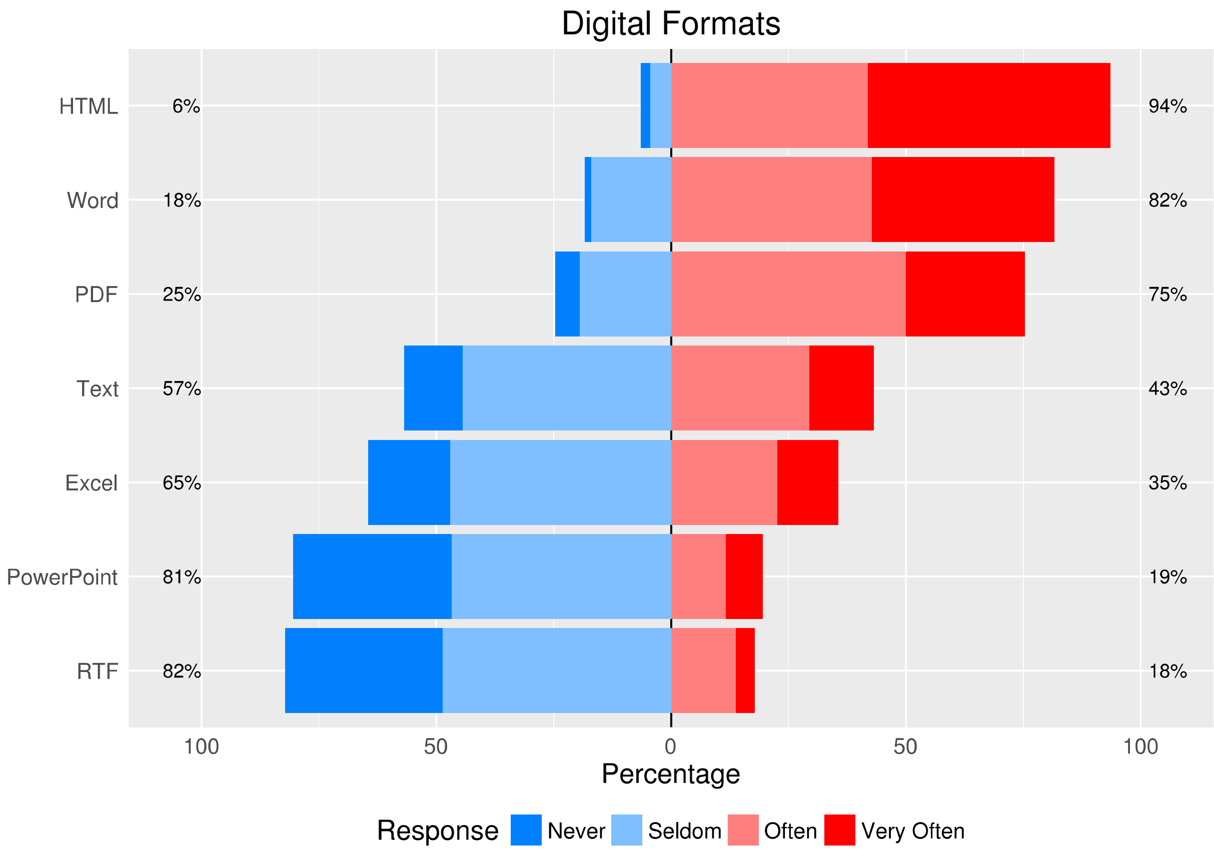


Figure 4.12: Digital Formats in which Documents are provided to Research Subjects

To 94 per cent of blind and visually impaired persons structured documents are provided in the format of HTML websites often and very often and only 6 per cent of them receive HTML documents seldom and never. Because of the emerging category of web based services and platforms on which structured documents are provided to them, this is the digital content format most used by the respondents for structured documents.

Because of the strong market share of Microsoft Office Word as an authoring tool for structured documents both, in industry as well as in private home offices, structured documents are provided to 82 per cent of the persons concerned in that format often and very often and only 18 per cent of them receive structured documents in the Microsoft Word format seldom or never.

Often, structured documents can be downloaded or are attached to E-Mails in the Adobe Portable Document Format (PDF) because this format established itself as an ISO standard for the authoring tool independent exchange of paged structured documents on the Internet. 75 per cent of blind and visually impaired persons receive structured documents in the Adobe Portable Document Format (PDF) often and very often and only 25 per cent of them receive structured PDF documents seldom or never.

The advantage of the plain text only format is that the text is well suited for the reading with screen readers using speech synthesis or Braille display output. On the other hand, the logical document structure is not semantically contained in the document which is a big disadvantage. In the plain text format structured documents are provided to 43 per cent of the persons concerned often and very often and 57 per cent of them receive plain text structured documents seldom and never.

Spreadsheets in the Microsoft Office Excel format are provided to 35 per cent of blind and visually impaired persons often and very often and 65 percent of the persons concerned receive Microsoft Office Excel spreadsheet seldom and never.

Slide show presentations in the Microsoft Office PowerPoint format are provided to 81 per cent of the persons concerned seldom and never and only 19 percent of the respondents receive Microsoft Office PowerPoint presentations often or very often.

To only 18 per cent of blind and visually impaired persons structured documents are provided in the Rich Text Format (RTF), a format proposed by Microsoft which allows the storage of some basic structural information in addition to the text, and 82 per cent of them receive structured documents in the RTF format seldom or never.

In contrast to the current situation, most blind and visually impaired persons would wish to receive their structured documents in the Microsoft Office Word format. Figure 4.13 shows the different digital content formats reported by blind and visually impaired persons in which they would wish to receive their structured documents as well as the frequencies and percentages of respondents in each category of digital content format wished:

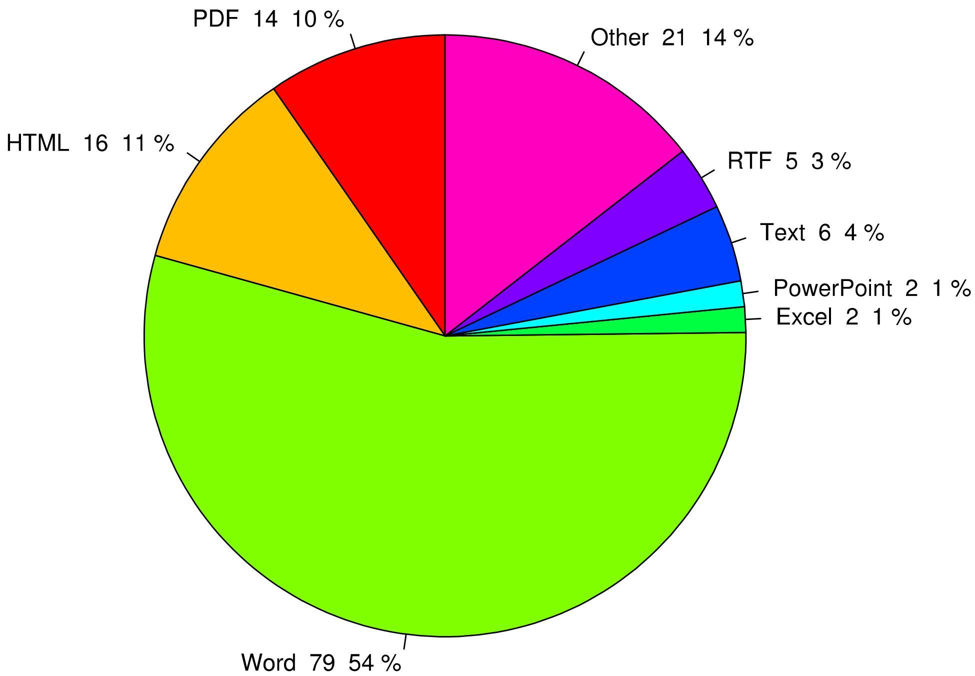


Figure 4.13: Digital Formats Research Subjects would wish to receive their Documents

54 per cent of blind and visually impaired persons would wish to receive their structured documents in the Microsoft Office Word format because this format is directly accessible to a variety of different assistive technologies and can easily be adapted to the special needs, for example by changing the font type, size or colour. 11 per cent of the persons concerned would wish to receive their structured documents as HTML websites. 10 per cent of the respondents would wish that structured documents are provided to them in the Adobe Portable Document Format (PDF), 4 per cent as plain text, 1 per cent as Microsoft Office Excel spreadsheet, another 1 per cent as Microsoft Office PowerPoint slide show presentation and 14 per cent of the participants would wish to receive their structured documents in other digital content formats.

Therefore a novel concept for an assistive technology in support of blind and visually impaired persons for the reading of structured documents should be format independent in order to be able to interworking with all the different digital content formats available out there because structured documents are provided to blind and visually impaired persons in a mix of different digital content formats.

4.3.4 Assistive Technologies

Blind and visually impaired persons are using different assistive technologies for the reading of structured documents. In most cases, a combination of multiple different assistive technologies employing different input and output modalities are used simultaneously at the same time in order to serve the special needs on different human senses like the visual, the aural or the tactile sense. These different assistive technologies are described in more detail in Chapter 2.2.1.

For desktop computers a combination of screen reader, screen magnifier with speech synthesis and Braille displays are used. Figure 4.14 shows the different assistive technologies reported by the persons concerned, which they use for the reading of structured documents on desktop computers as well as the percentages of how often each assistive technology is used:

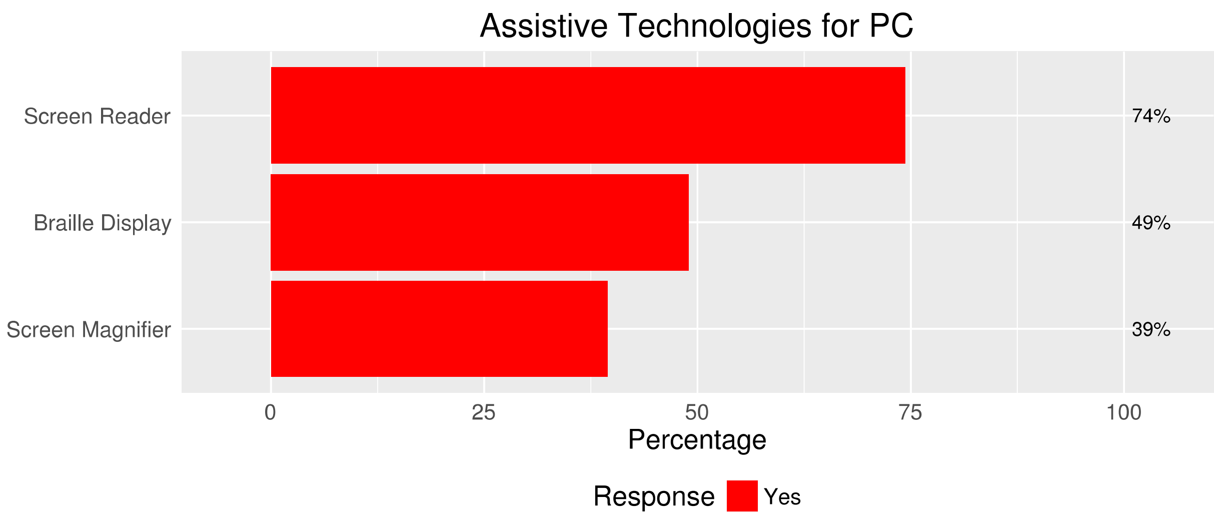


Figure 4.14: Assistive Technologies used by Research Subjects for PC

74 per cent of blind and visually impaired persons are using a screen reader on desktop computers for the reading of structured documents.

The use of screen readers for desktop computers depends to 34 per cent on the visual impairment category a specific person is in, as shown by a moderate bivariate correlation with v = 0.59 and a very high level of statistical significance with p = 1.52e-8. Figure 4.15 shows the use of screen readers for desktop computers in correlation to the different categories of visual impairment as well as the percentages of how often they are used by each visual impairment category:

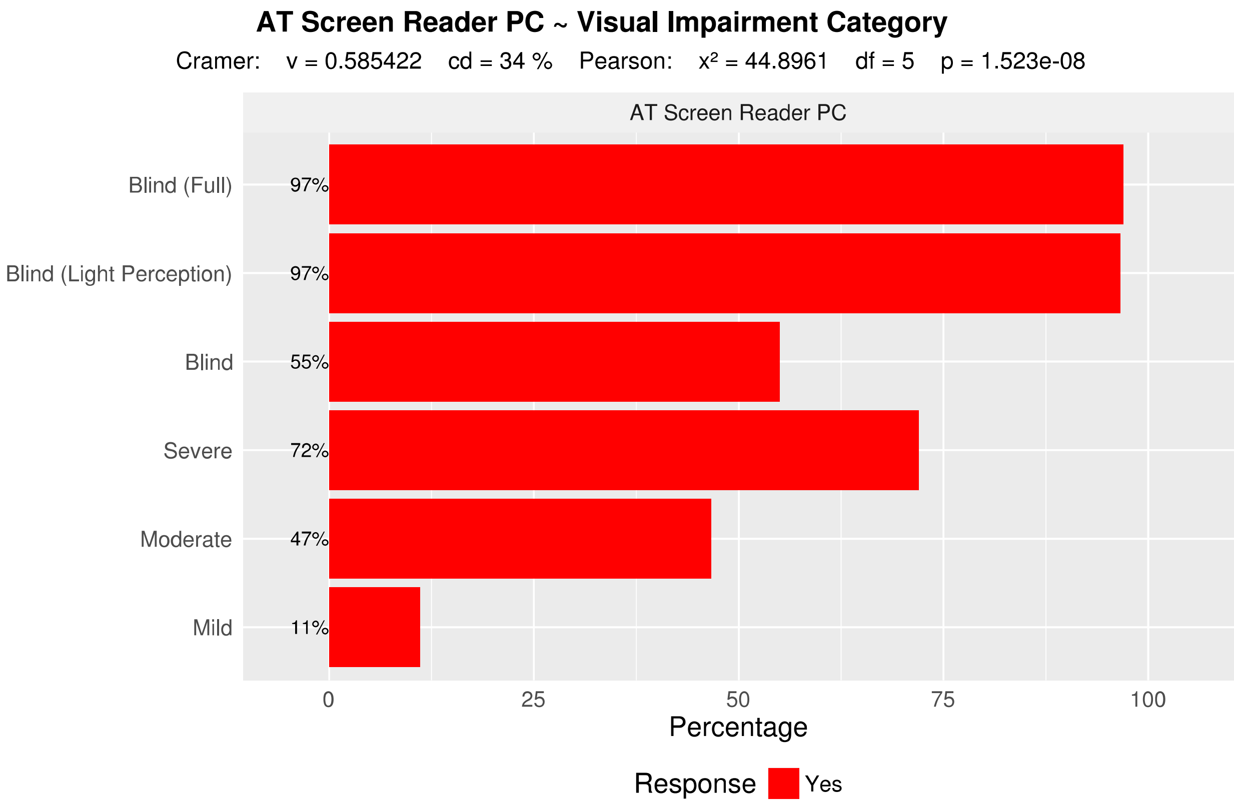


Figure 4.15: Screen Readers for PC correlated to Visual Impairment Categories

97 per cent of fully blind persons without light perception as well as another 97 percent of blind persons with light perception, 55 per cent of blind persons with a visus worse than 0.05, 72 per cent of persons with a severe visual impairment, 47 per cent of persons with a moderate visual impairment and only 11 per cent of persons with a mild visual impairment are using a screen reader for desktop computers. The categories of persons with low vision typically do not use screen readers for desktop computers because their visual performance is sufficient for reading structured documents visually using screen magnification or even without using any assistive technologies at all.

49 per cent of the persons concerned are using a Braille display for the reading of structured documents on desktop computers.

The use of Braille displays for desktop computers is, as with the use of screen readers described above, to 66 per cent dependent on the visual impairment category a specific person is in, as shown by a very strong bivariate correlation with v = 0.81 and a very high statistical significance level of p = 2.96e-17. Figure 4.16 shows the use of Braille displays for desktop computers in correlation to the different categories of visual impairment as well as the percentages of how often they are used by each visual impairment category:

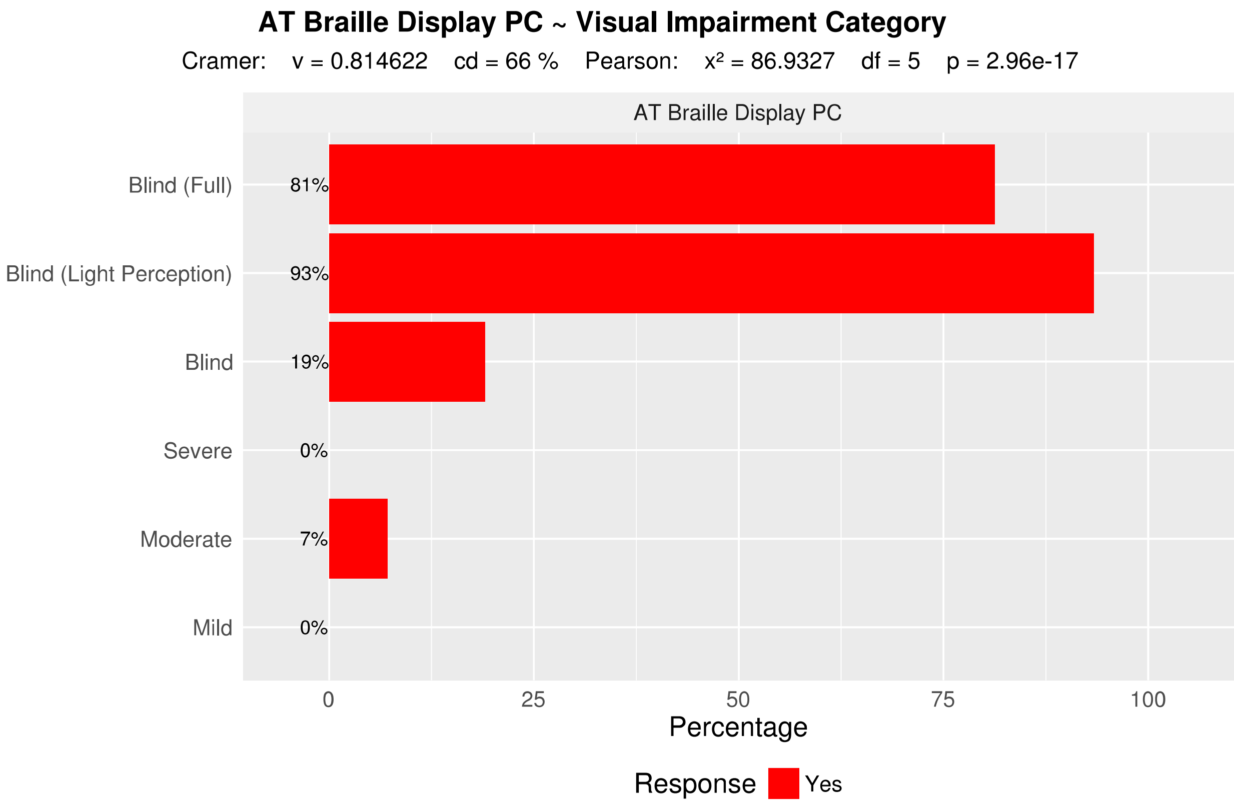


Figure 4.16: Braille Displays for PC in correlation to Visual Impairment Categories

81 per cent of fully blind persons without light perception as well as 93 per cent of blind persons with light perception, 19 per cent of blind persons with a visus worse than 0.05 and only 7 percent of persons with a severe visual impairment are using a Braille display for desktop computers. The categories of persons with a severe visual impairment and persons with a mild visual impairment do not use Braille displays for desktop computers at all because their visual performance is sufficient for reading structured documents without this extra assistive technology.

Another strong bivariate correlation with v = 0.61 and a very high statistical significance level of p = 7.61e-12 shows that the use of Braille displays for desktop computers also depends to 38 per cent on the use of the embossed physical medium. Figure 4.17 shows the use of Braille displays for desktop computers in correlation to the use of the embossed physical medium as well as the percentages of how often they are used:

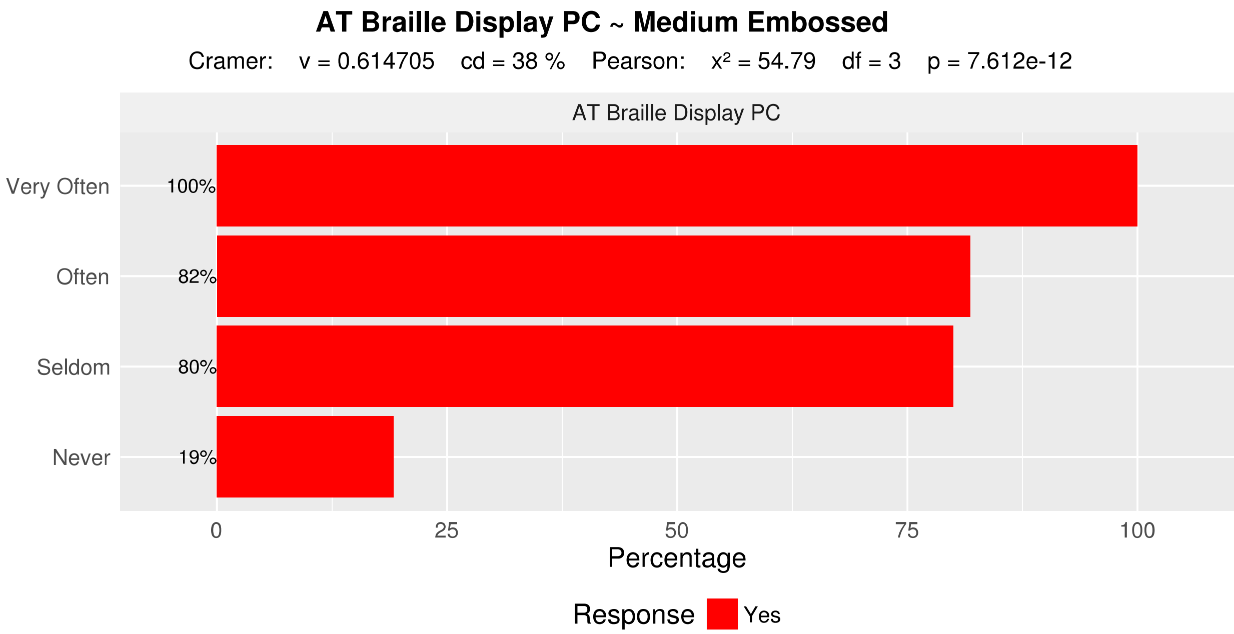


Figure 4.17: Braille Displays for PC in correlation to Physical Medium Embossed

100 per cent of the persons who receive structured documents very often embossed on paper, 82 per cent of persons receiving embossed structured documents often as well as 80 per cent of persons who receive structured documents embossed on paper seldom and only 19 per cent of the persons never receiving embossed structured documents are also using a Braille display for desktop computers.

Another moderate bivariate correlation witch v = 0.42 and a very high statistical significance level of p = 1.28e-7 shows that the use of Braille displays for desktop computers also depends to 18 per cent on the use of screen readers. Figure 4.18 shows the use of Braille displays for desktop computers in correlation to the use of screen readers for desktop computers as well as the percentages of how often Braille displays are used:

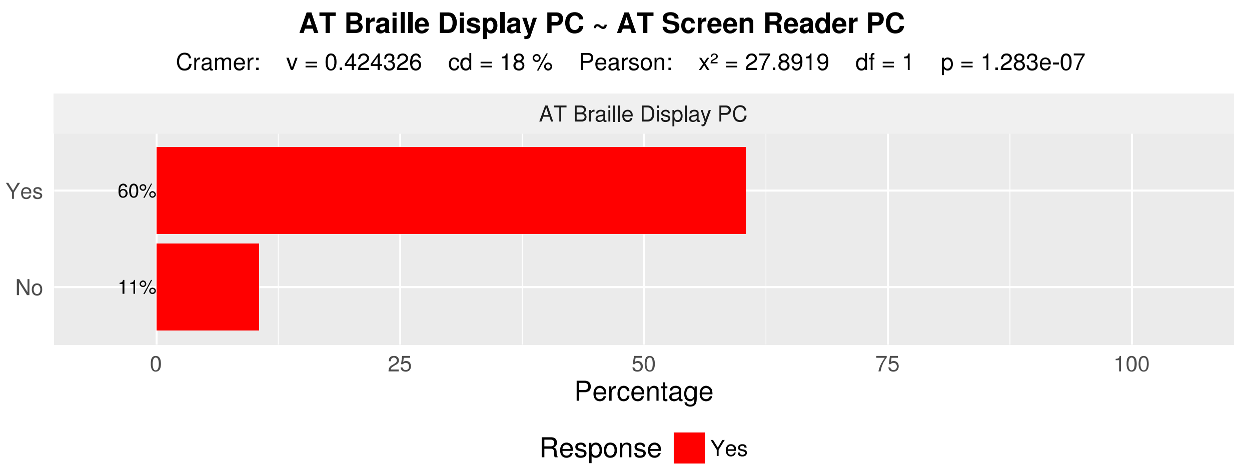


Figure 4.18: Braille Displays for PC in correlation to Screen Readers for PC

60 per cent of the persons who are using screen readers on desktop computers and only 11 per cent of persons not using screen readers for desktop computer are also using a Braille display for desktop computers.

Only 39 per cent of blind and visually impaired persons are using a screen magnifier for reading structured documents on desktop computers. A screen magnifier gets the screen content by the computer's graphics output interface and presents it to the user in an enlarged form by using different magnification ranges and magnification modes like full screen, lens or fixed magnified portion. Additional features like colour inversion, font enhancements like text smoothing, focus enhancements, cursor customisation, control keys and mouse wheel to zooming in and out may be provided. It supports persons with low vision or deteriorating sight.

As with the use of screen readers and Braille displays described above, the use of screen magnifiers for desktop computers depends to 47 per cent on the visual impairment category a specific person is in, as shown by a strong bivariate correlation with v = 0.69 and a very high level of statistical significance with p = 1.16e-11. Figure 4.19 shows the use of screen magnifiers for desktop computers in correlation to the different categories of visual impairment as well as the percentages of how often screen magnifiers are used by each visual impairment category:

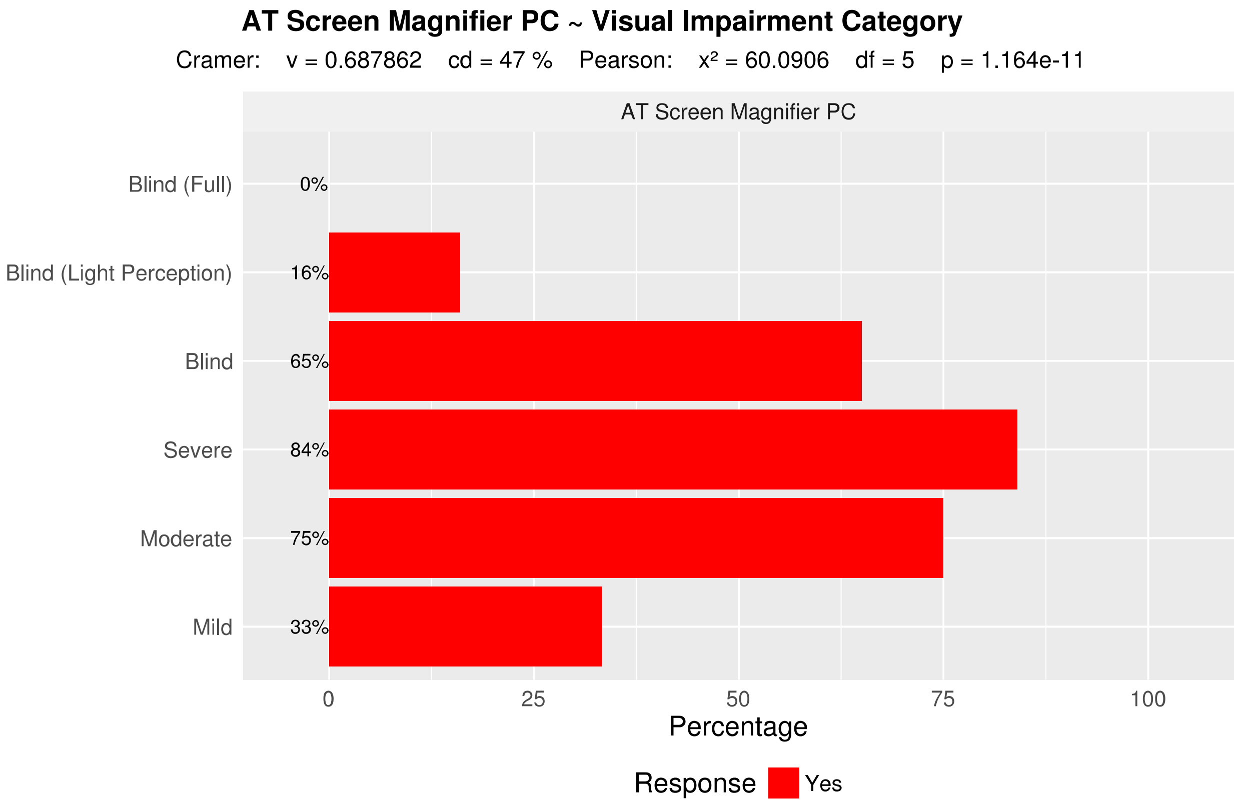


Figure 4.19: Screen Magnifiers for PC in correlation to Visual Impairment Categories

33 per cent of persons with a mild visual impairment as well as 75 per cent of persons with a moderate visual impairment, 84 per cent of persons with a severe visual impairment, 65 per cent of blind persons with a visus worse than 0.05 and only 16 per cent of blind persons with light perception are using a screen magnifier for desktop computers. The category of fully blind persons without light perception do not use screen magnifiers at all because this visual assistive technology does not provide any additional value for them. Also persons with a mild visual impairment typically do not use a screen magnifier because their visual performance is sufficient for reading structured documents on desktop computers without using any assistive technologies at all.

Another moderate bivariate correlation with v = 0.47 and a very high statistical significance level of p = 1.76e-8 shows that the use of screen magnifiers for desktop computers depends to 22 per cent on the use of Braille displays. Figure 4.20 shows the use of screen readers for desktop computers in correlation to the user of Braille displays for desktop computers as well as the percentages of how often they are used:

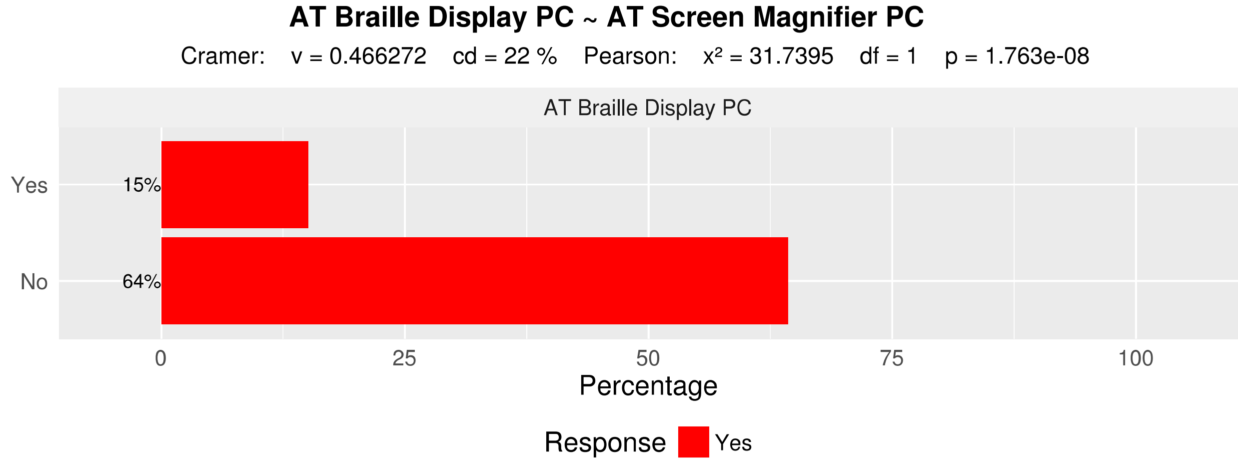


Figure 4.20: Screen Magnifiers for PC in correlation to Braille Displays

85 per cent of persons who do not use a Braille display for desktop computers and only 15 per cent of persons who use a Braille display for desktop computer are also using a screen magnifier in addition.

The assistive technology products available on the market can either be proprietary or built into the operating system. They can be categorised into commercial as well as in free or open source solutions.

The screen reader products used by blind and visually impaired persons for the reading of structured documents on desktop computers is a mix of different products too. Figure 4.21 shows the different screen reader products for desktop computers reported by the persons concerned as well as the frequencies and percentages of respondents using each screen reader product:

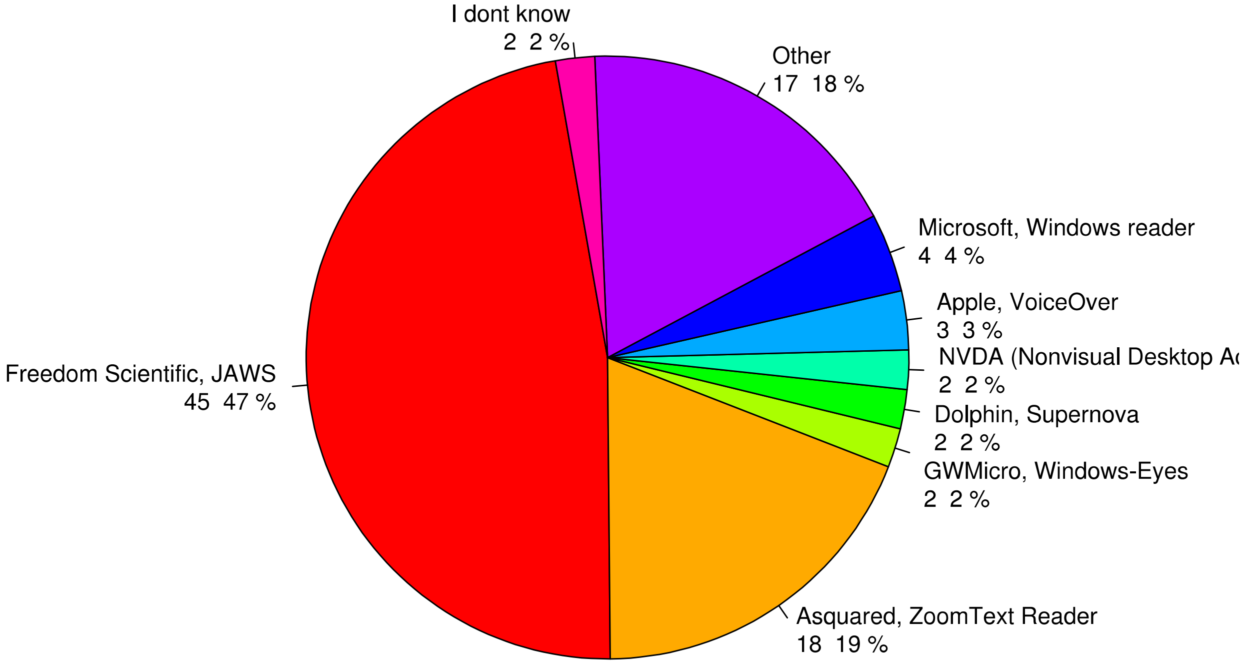


Figure 4.21: Screen Reader Products used by Research Subjects for PC

47 per cent of blind and visually impaired persons are using Freedom Scientific JAWS [JAWS], a commercial screen reader for Microsoft Windows which can be used in combination with the MAGic screen magnifier [MAGic] from the same company. This is the most common used product because of the large market share of Microsoft Windows as an operating system for desktop computers.

Aisquared ZoomText Reader [ZoomTextReader], a basic commercial screen reader for the Microsoft Windows operating system too, optimised for interworking with the ZoomText magnifier [ZoomText] screen magnifier from the same manufacturer, is used by 19 per cent of the persons concerned.

4 per cent of blind and visually impaired persons are using Microsoft Windows Reader [WindowsReader], a very basic screen reader that comes with the Microsoft Windows operating system free of charge.

3 per cent of the persons concerned are using Apple VoiceOver [VoiceOver], the screen reader included free of charge in Apple's MacOS operating system.

Dolphin Supernova [Supernova], a commercial screen reader for the Microsoft Windows operating system, designed for use in combination with the Dolphin Lunar screen magnifier [Lunar] manufactured by the same company, is used by 2 per cent of the persons concerned.

Another 2 per cent of blind and visually impaired persons are using Nonvisual Desktop Access NVDA [NVDA], a free and open source screen reader project for the Microsoft Windows platform.

GWMicro Windows Eyes [WindowsEyes], a commercial screen reader for the Microsoft Windows operating system, is very frequently used in the United State, but in Europe it is only used by 2 per cent of the respondents.

Finally, 18 per cent of blind and visually impaired persons are using another screen reader product and 2 per cent of the participants reported not to knowing which screen reader product they are using.

The screen magnifier products used by blind and visually impaired persons for the reading of structured documents on desktop computers is, as with the screen reader products described above, a mix of different products. Figure 4.22 shows the different screen magnifier products for desktop computers reported by the persons concerned as well as the frequencies and percentages of respondents using each screen magnifier product:

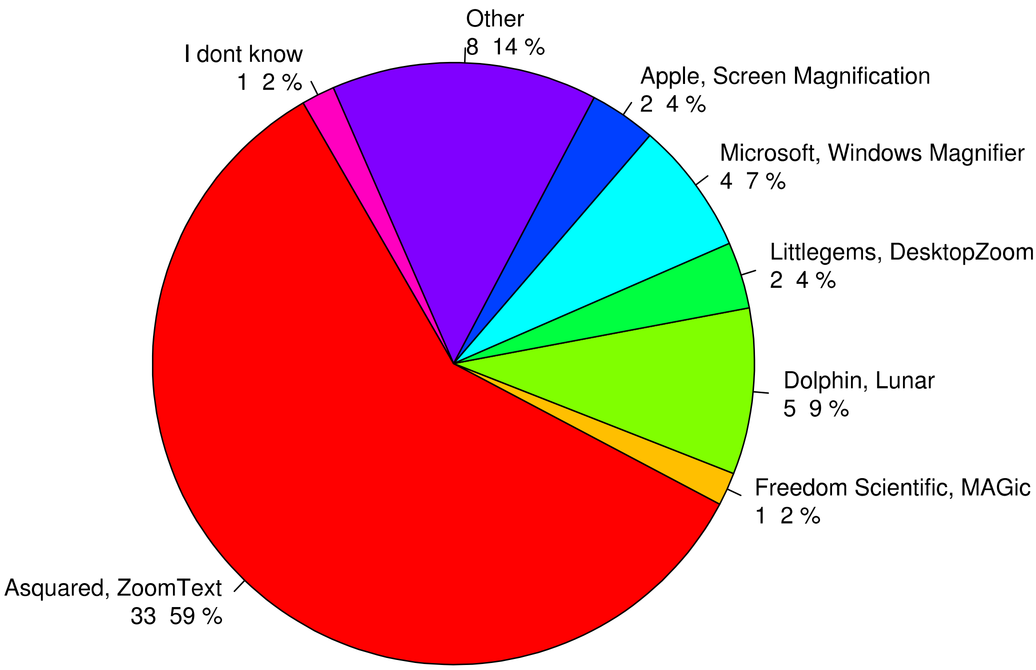


Figure 4.22: Screen Magnifier Products used by Research Subjects for PC

59 per cent of blind and visually impaired persons are using Aisquared ZoomText [ZoomText], a commercial screen magnifier for the Microsoft Windows platform. This is the most common used product because of the huge market share of Microsoft Windows as an operating system for desktop computers.

Dolphin Lunar [Lunar], a commercially available screen magnifier for the Microsoft Windows operating system which can be used in combination with the Dolphin Supernova [Supernova] screen reader from the same company, is used by 9 per cent of the persons concerned.

7 per cent of blind and visually impaired persons use Microsoft Windows Magnifier [WindowsMagnifier], a basic screen magnifier coming with the Microsoft Windows operating system free of charge.

Freedom Scientific MAGic [MAGic], as with ZoomText and Lunar a commercial screen magnifier product for the Microsoft Windows operating system and designed for use in combination with the JAWS screen reader [JAWS] manufactured by the same company, is used by 2 per cent of the persons concerned.

4 per cent of blind and visually impaired persons are using Apple Zoom [AppleZoom], the screen magnifier included free of charge in Apple's MacOS operation system.

Littlegems DesktopZoom [DesktopZoom], a very considerable free screen magnifier for the Microsoft Windows operating system, is used by another 4 per cent of the persons concerned.

Finally, 14 per cent of the respondents are using another screen magnifier product and only 2 per cent of the participants reported not to knowing which screen magnifier product they are using.

In the category of mobile and wearable devices like smart phones and smart tablets, a combination of screen reader and screen magnifier are used. At the moment, there are no Braille displays available for mobile and wearable devices. Figure 4.23 shows the different assistive technologies reported by the persons concerned, which they use for the reading of structured documents on mobile and wearable devices as well as the percentages of how often each assistive technology is used:

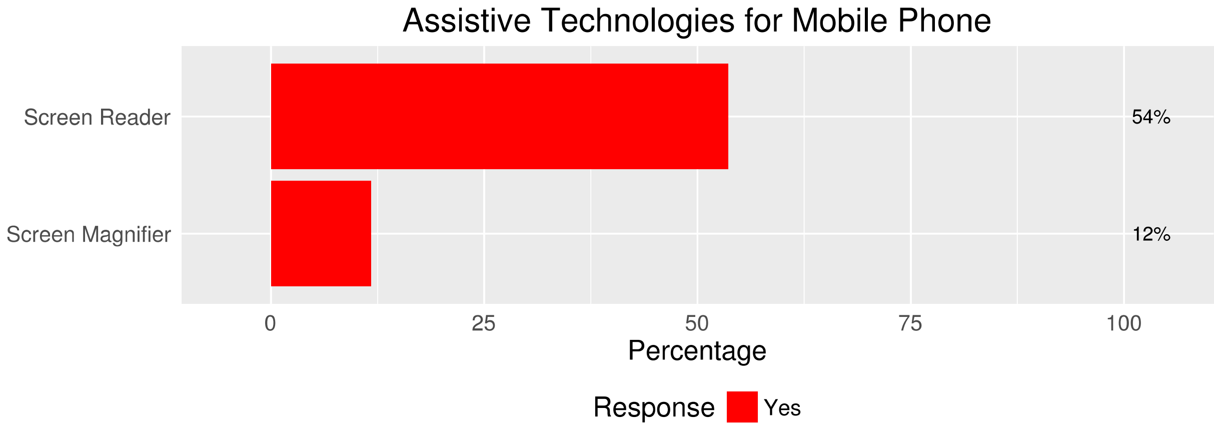


Figure 4.23: Mobile Assistive Technologies used by the Persons Concerned

54 per cent of the blind and visually impaired persons are using a screen reader on mobile and wearable devices for the reading of structured documents.

The use of screen readers on mobile and wearable devices depends to 28 per cent on the visual impairment category a specific person is in, as shown by a moderate bivariate correlation with v = 0.53 and a very high level of statistical significance with p = 1.01e-6. Figure 4.24 shows the use of screen readers on mobile and wearable devices in correlation to the different categories of visual impairment as well as the percentages of how often they are used by each visual impairment category:

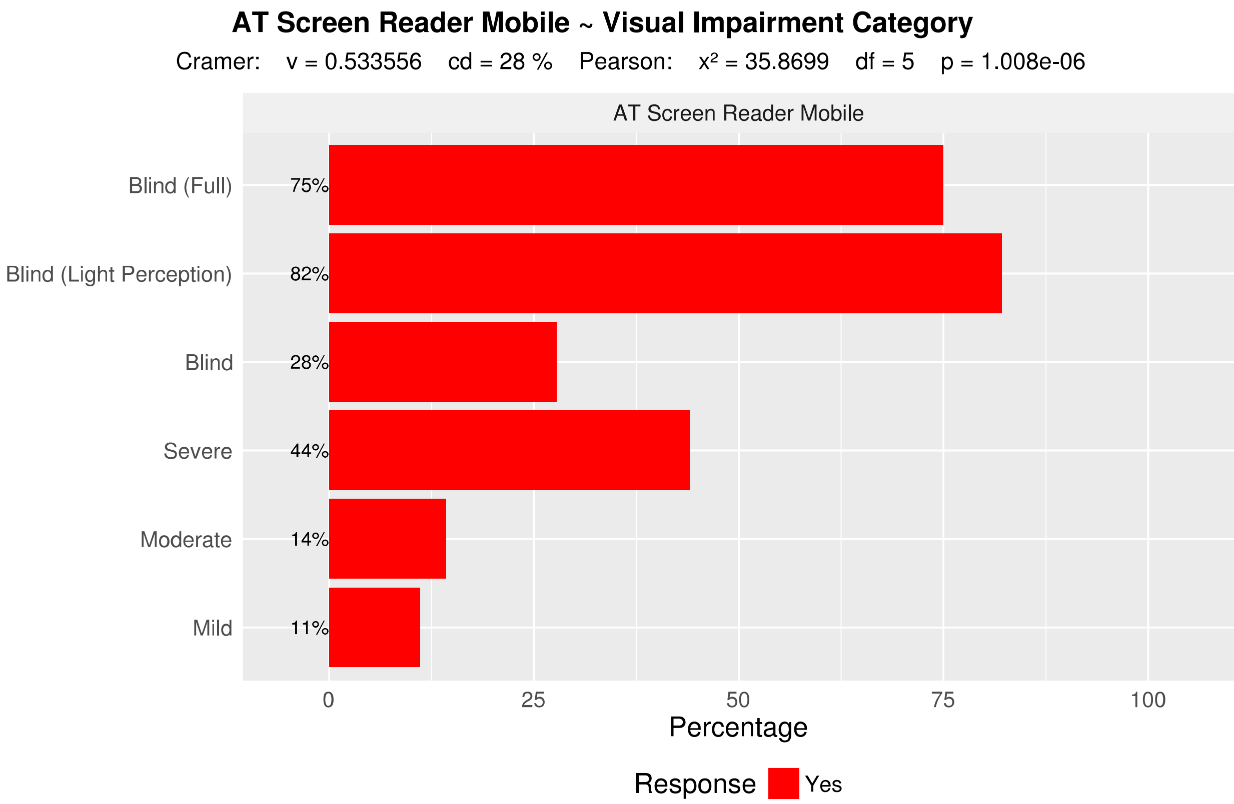


Figure 4.24: Mobile Screen Readers in correlation to Visual Impairment Categories

75 per cent of fully blind persons without light perception as well as 82 percent of blind persons with light perception, 28 per cent of blind persons with a visus worse than 0.05,

44 per cent of persons with a severe visual impairment, 14 per cent of persons with a moderate visual impairment and only 11 per cent of persons with a mild visual impairment are using a screen reader on mobile and wearable devices. The categories of persons with low vision typically do not use screen readers on mobile and wearable devices because their visual performance is sufficient for reading structured documents visually using screen magnification or even without using any assistive technology at all.

Another moderate bivariate correlation with v = 0.39 and a very high statistical significance level of p = 1.39e-6 shows that the use of screen readers on mobile and wearable devices depends to 15 per cent on the use of screen readers for desktop computers. Figure 4.25 shows the use of screen readers on mobile and wearable devices in correlation to the use of screen readers for desktop computers as well as the percentages of how often they are used:

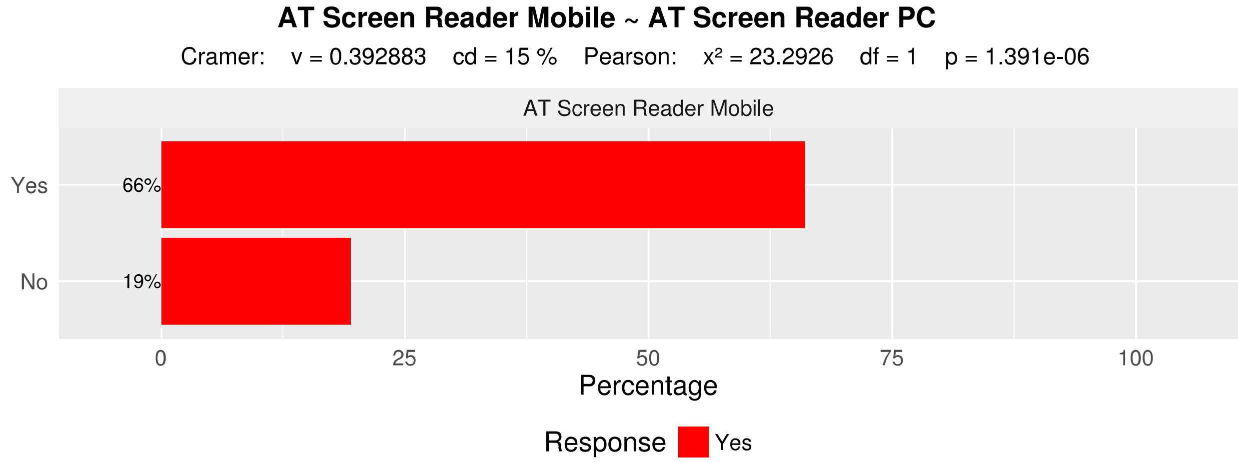


Figure 4.25: Mobile Screen Readers in correlation to Screen Readers for PC

66 per cent of the persons who are using screen readers for desktop computers and only 19 per cent of persons not using screen readers for desktop computers do also use a screen reader on mobile and wearable devices.

Another moderate bivariate correlation with v = 0.47 and a very high statistical significance level of p = 1.38e-8 shows that the use of screen readers on mobile and wearable devices depends to 22 per cent on the use of Braille displays for desktop computers. Figure 4.26 shows the use of screen readers on mobile and wearable devices in correlation to the use of Braille displays for desktop computers as well as the percentages of how often they are used:

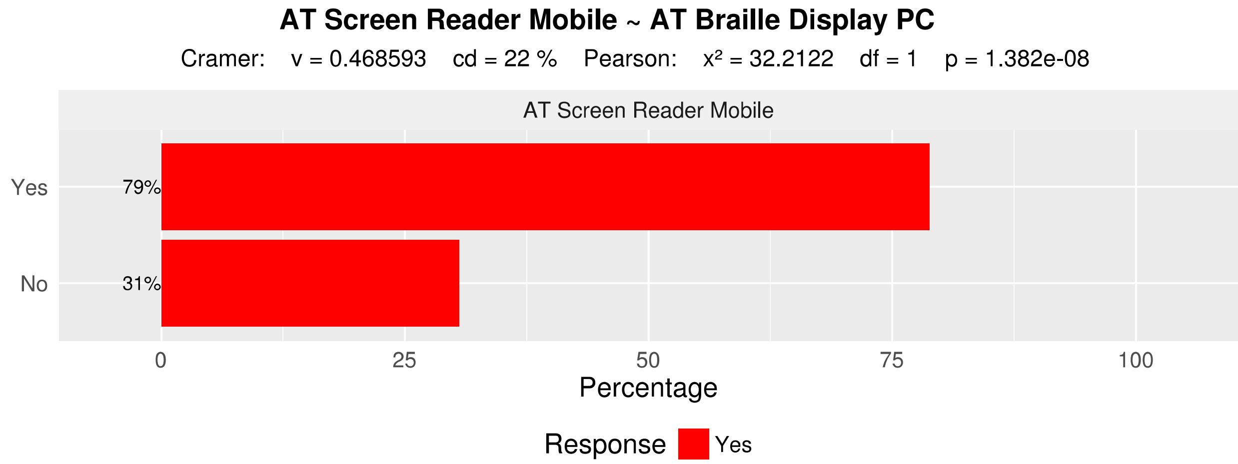


Figure 4.26: Mobile Screen Readers in correlation to Braille Displays for PC

79 per cent of the persons who are using Braille displays for desktop computers and only 31 per cent of the persons not using Braille displays for desktop computers are also using a screen reader on mobile and wearable devices in addition.

Only 12 per cent of the blind and visually impaired persons are using a screen magnifier on mobile and wearable devices for the reading of structured documents.

As with the use of screen readers described above, the use of screen magnifiers on mobile and wearable devices depends to 10 per cent on the visual impairment category a specific person is in, as shown by a weak bivariate correlation with v = 0.32 and a high level of statistical significance with p = 0.029. Figure 4.27 shows the use of screen magnifiers for desktop computers in correlation to the different categories of visual impairment as well as the percentages of how often screen magnifiers are used by each visual impairment category:

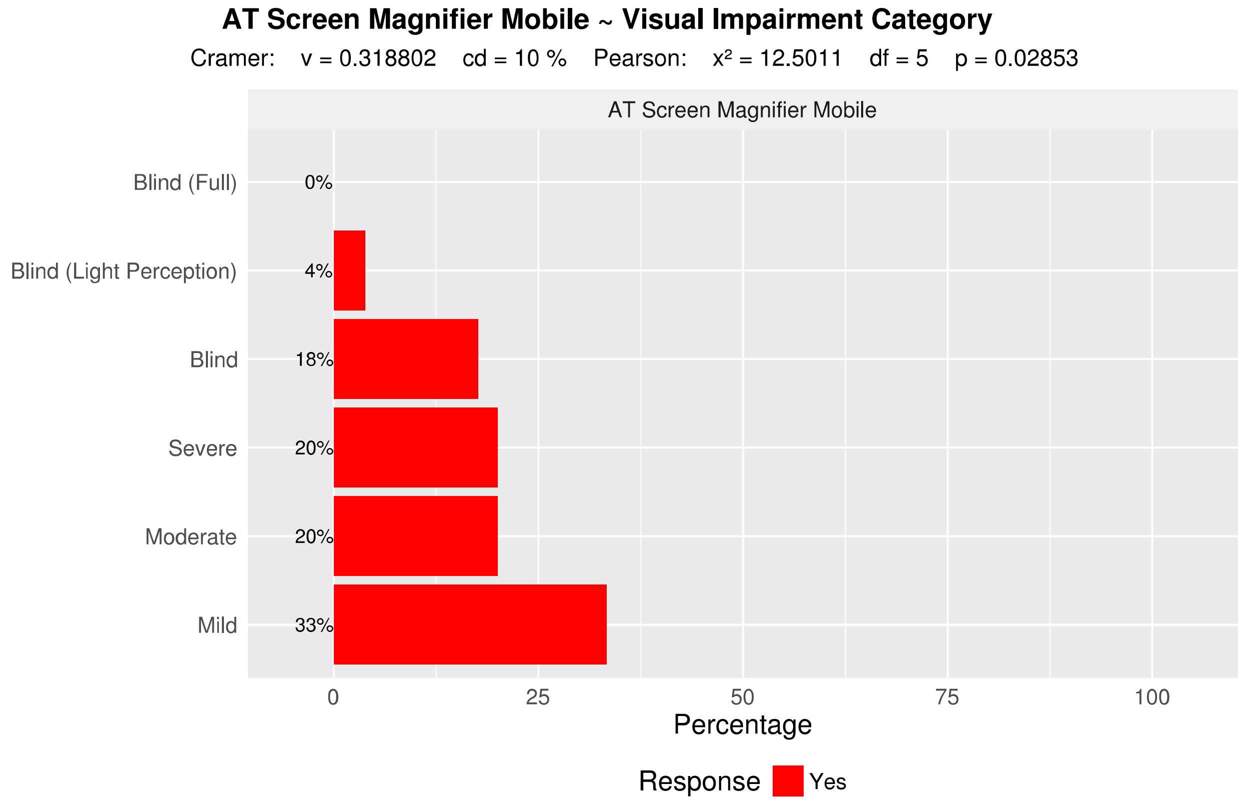


Figure 4.27: Mobile Screen Magnifiers in correlation to Visual Impairment Categories

33 per cent of persons with a mild visual impairment as well as 20 per cent of persons with a moderate visual impairment, another 20 per cent of persons with a severe visual impairment, 18 per cent of blind persons with a visus worse than 0.05 and only 4 per cent of blind persons with light perception are using a screen magnifier on mobile and wearable devices. The category of fully blind persons without light perception do not use screen magnifiers at all because this visual assistive technology does not provide any additional value for them.

The screen reader products used by blind and visually impaired persons for the reading of structured documents on mobile and wearable devices is a mix of different products too. Figure 4.28 shows the different screen reader products for mobile and wearable devices reported by the persons concerned as well as the frequencies and percentages of respondents using each screen reader product:

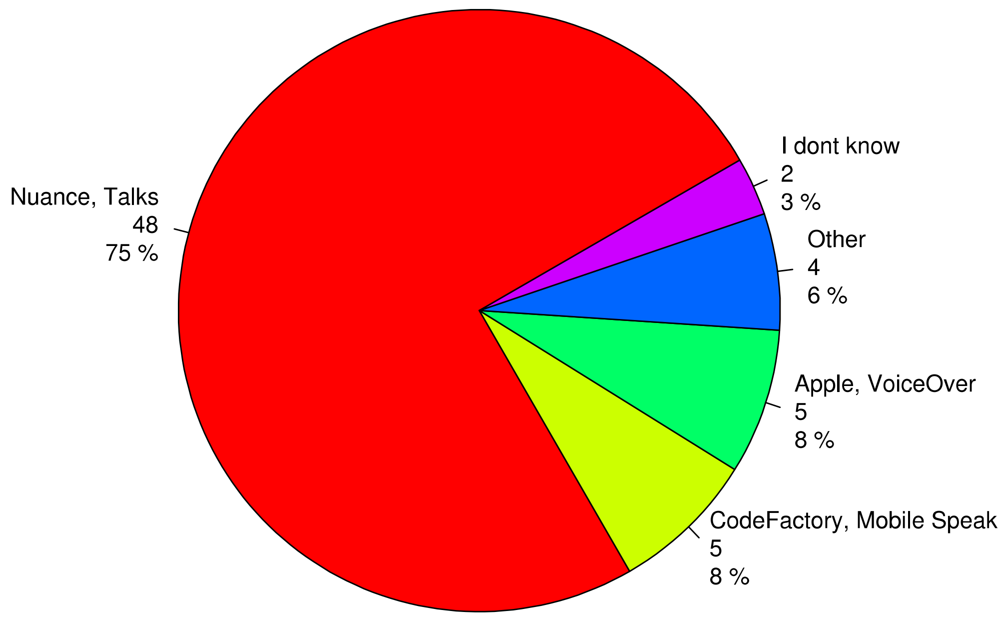


Figure 4.28: Mobile Screen Reader Products used by the Persons Concerned

75 per cent of the blind and visually impaired persons are using Nuance Talks [Talks], a commercial screen reader for Nokia's Symbian operating system, because of the large market share of Nokia mobile devices. In the future the market share of this product may decrease in favour of the emerging categories of Apple iOS [iOS] and Google Android [Android] mobile and wearable devices.

CodeFactory Mobile Speak [MobileSpeak], a commercial screen reader available for both, the Nokia SymbianOS as well as for the Microsoft Windows Mobile operating system, is used by 8 per cent of the persons concerned.

8 per cent of the respondents use Apple VoiceOver for iOS [VoiceOverIOS], the screen reader included in Apple's iOS operating system free of charge. In the future, the market share of this product may increase because Apple's iDevices are very popular with blind and visually impaired persons due to the fact that Apple was the first manufacturer, providing assistive technologies like screen reader and screen magnifier out of the box and free of charge.

6 per cent of the blind and visually impaired persons are using another mobile screen reader and 3 per cent of the participants reported to not knowing which mobile screen reader product they are using.

The screen reader products used on mobile and wearable devices depend to 18 per cent on the use of screen magnifiers, as shown by a moderate bivariate correlation with $v = 0.42$ and a high statistical significance level of $p = 0.039$. Figure 4.29 shows the screen reader products used on mobile and wearable devices in correlation to the use of screen magnifiers on mobile and wearable devices as well as the frequencies and percentages of how often each screen reader product is used:

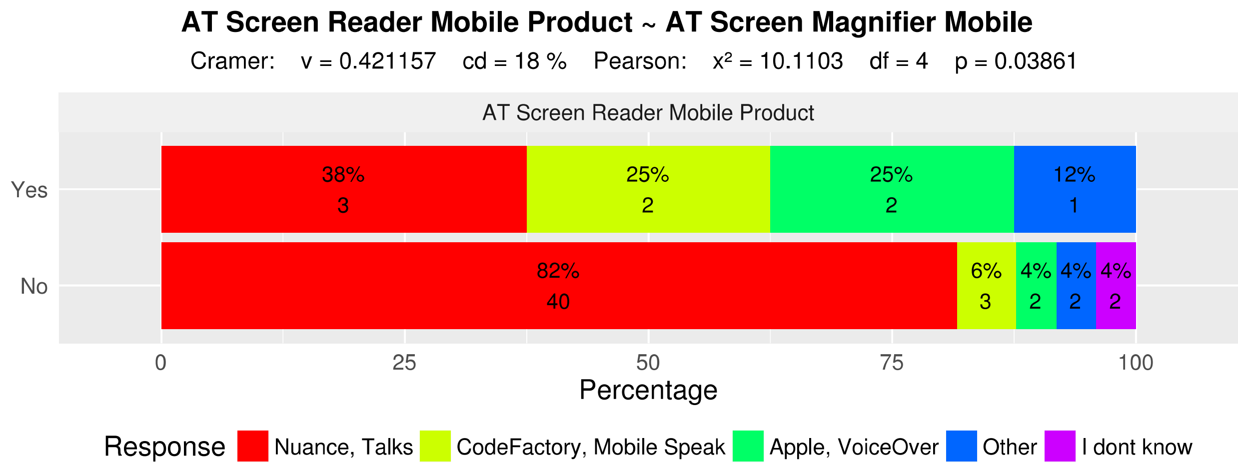


Figure 4.29: Mobile Screen Reader Products to Mobile Screen Magnifiers

38 per cent of persons using a screen magnifier and 82 per cent of persons who are not using a screen magnifier use Nuance Talks. CodeFactory Mobile Speak is used by 25 per cent of the persons using a screen magnifier and only 6 per cent of persons who do not use a screen magnifier. Another 25 per cent of persons using a screen magnifier and only 4 per cent of persons who do not use a screen magnifier are using Apple VoiceOver. Another screen reader product is used by 12 per cent of persons using a screen magnifier and by only 4 per cent of persons who do not use a screen magnifier. Finally, another 4 per cent of persons who do not use a screen magnifier reported to not knowing which screen reader product they are using. This correlation arises because not all combinations of screen magnifier and screen reader products are able to interworking together well and can be used simultaneously at the same time on a mobile and wearable device.

The screen magnifier products used by blind and visually impaired persons for the reading of structured documents on mobile and wearable devices is, as with the screen reader products described above, a mix of different products. Figure 4.30 shows the different screen magnifier products for mobile and wearable devices reported by the persons concerned as well as the frequencies and percentages of respondents using each screen magnifier product:

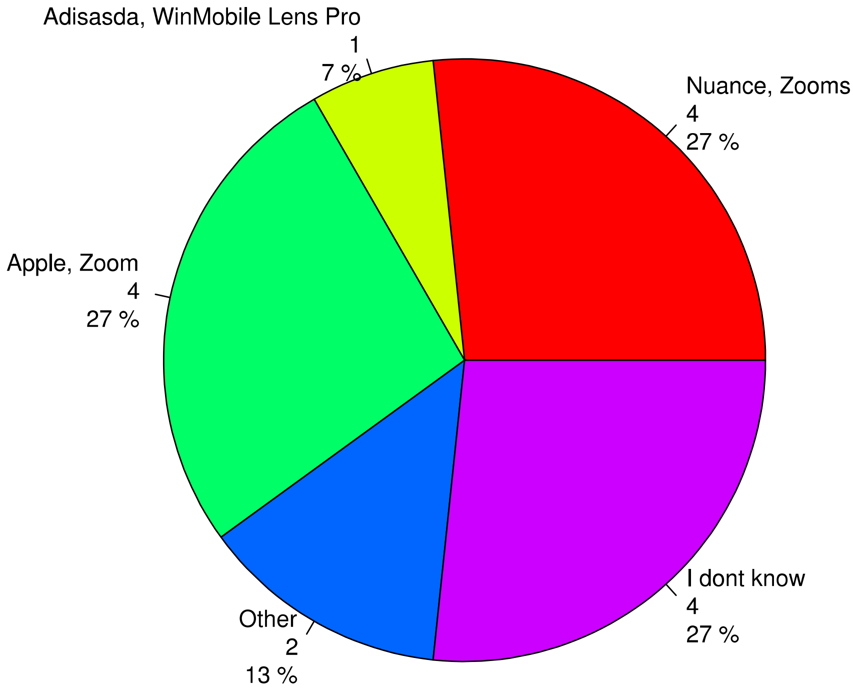


Figure 4.30: Mobile Screen Magnifier Products used by the Persons Concerned

27 per cent of blind and visually impaired persons are using Apple Zoom for iOS [AppleZoomIOS], the screen magnifier coming with Apple's iOS operating system free of charge, because Apple's iDevices are very popular with blind and visually impaired persons due to the fact that Apple was the first manufacturer, providing assistive technologies like screen reader and screen magnifier out of the box and free of charge.

Nuance Zooms [Zooms], a commercial screen magnifier for Nokia's Symbian operating system, is used by another 27 per cent of the persons concerned.

7% of the respondents use Adisasda WinMobile Lens Pro [WinMobileLensPro], a commercial screen magnifier for the Microsoft Windows Mobile operating system.

13 per cent of the blind and visually impaired persons are using another screen magnifier product and 27% of the participants reported not to knowing which mobile screen magnifier product they are using.

The screen magnifier products used on mobile and wearable devices depend to 65 per cent on the use of screen readers, as shown by a very strong bivariate correlation with $v = 0.81$ and a high statistical significance level of $p = 0.044$. Figure 4.31 shows the screen magnifier products used on mobile and wearable devices in correlation to the use of screen readers on mobile and wearable devices as well as the frequencies and percentages of how often each screen magnifier product is used:

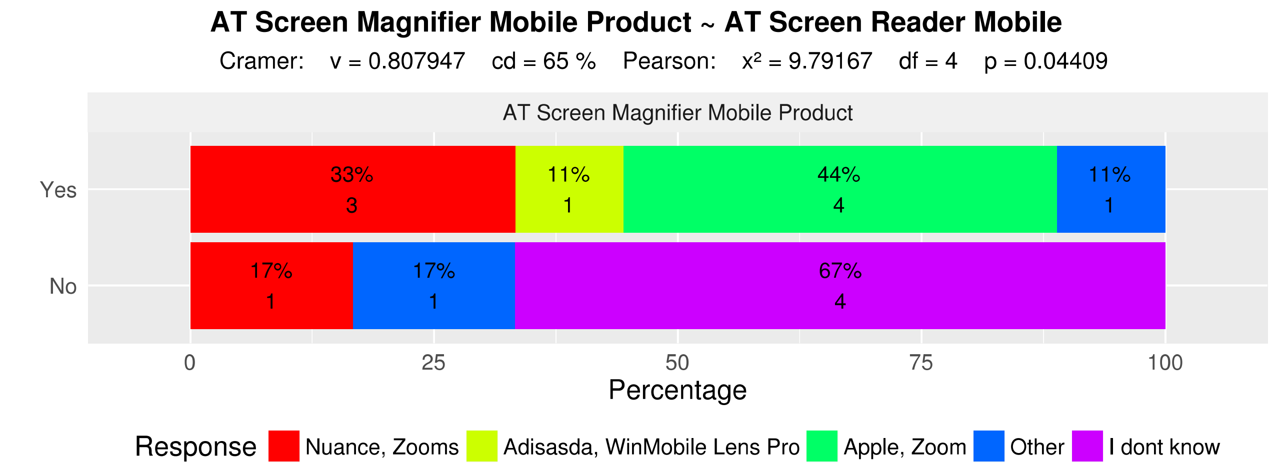


Figure 4.31: Mobile Screen Magnifier Products to Mobile Screen Readers

33 per cent of persons using a screen reader and 17 per cent of persons who are not using a screen reader use Nuance Zooms. Adisasda WinMobile Lens Pro is used by 11 per cent of the persons using a screen reader. 44 per cent of persons using a screen reader are using Apple Zoom. Another screen magnifier product is used by only 11 per cent of persons using a screen reader and by 17 per cent of persons who do not use a screen reader. Finally, 67 per cent of persons who do not use a screen reader reported to not knowing which screen magnifier product they are using. This correlation arises because not all combinations of screen reader and screen magnifier products are able to interworking together well and can be used simultaneously at the same time on a mobile and wearable device.

More than one human sense should be served by a novel concept for an assistive technology in support of blind and visually impaired persons for the reading of structured documents because in most cases, a combination of multiple assistive technologies employing different input and output modalities are used simultaneously at the same time in order to serve the special needs on multiple human senses like the visual, the aural or the tactile sense.

4.3.5 Input Modalities

Blind and visually impaired persons are using a combination of different input modalities for the reading of and the navigation within structured documents on desktop computers. Figure 4.32 shows the different input modalities reported by the persons concerned as well as the percentages of how often each input modality is used:

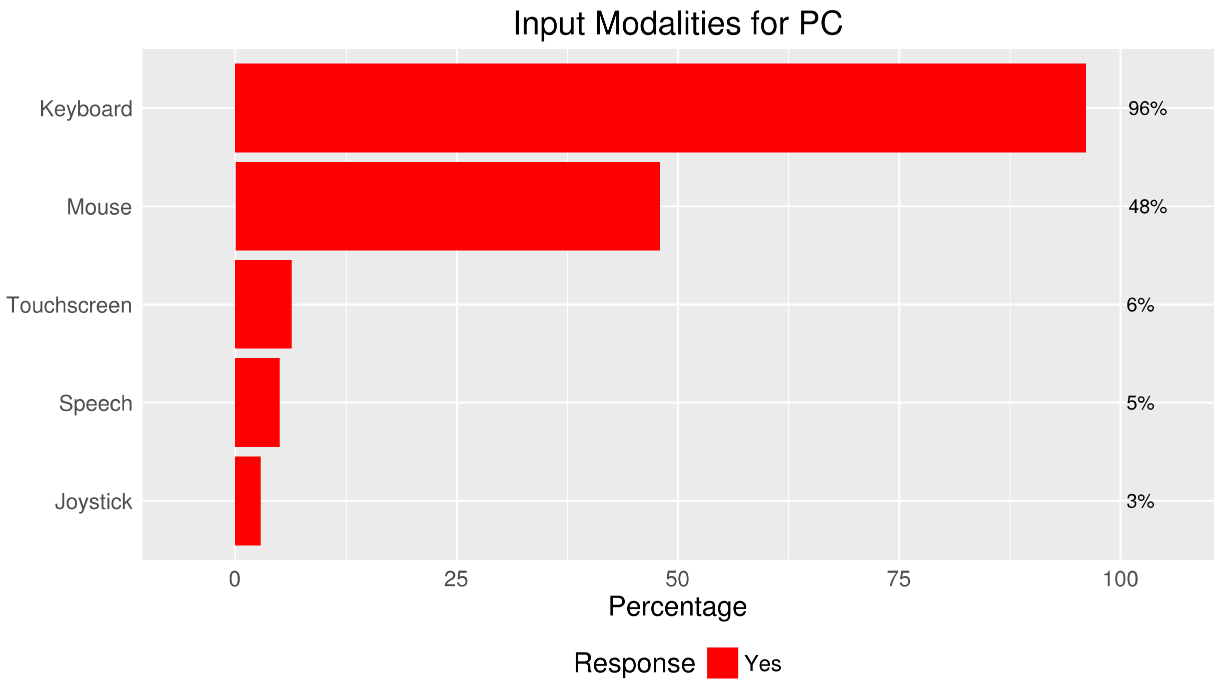


Figure 4.32: Input Modalities used for Navigation within Documents for PC

96 per cent of the blind and visually impaired persons are using keyboard for the reading of and the navigation within structured documents on desktop computers because most assistive technologies like screen reader or screen magnifier provide numerous key sequences for the navigation within structured documents, for example for jumping from one heading to the next.

48 per cent of the persons concerned are also using mouse for the reading of and the navigation within structured documents on desktop computers.

The use of mouse depends to 67 percent on the visual impairment category a specific person is in, as shown by a very strong bivariate correlation with v = 0.82 and very high level of statistical significance with p = 5.49e-17. Figure 4.33 shows the use of the mouse input modality in correlation to the different visual impairment categories as well as the percentages of how often mouse is used by each visual impairment category:

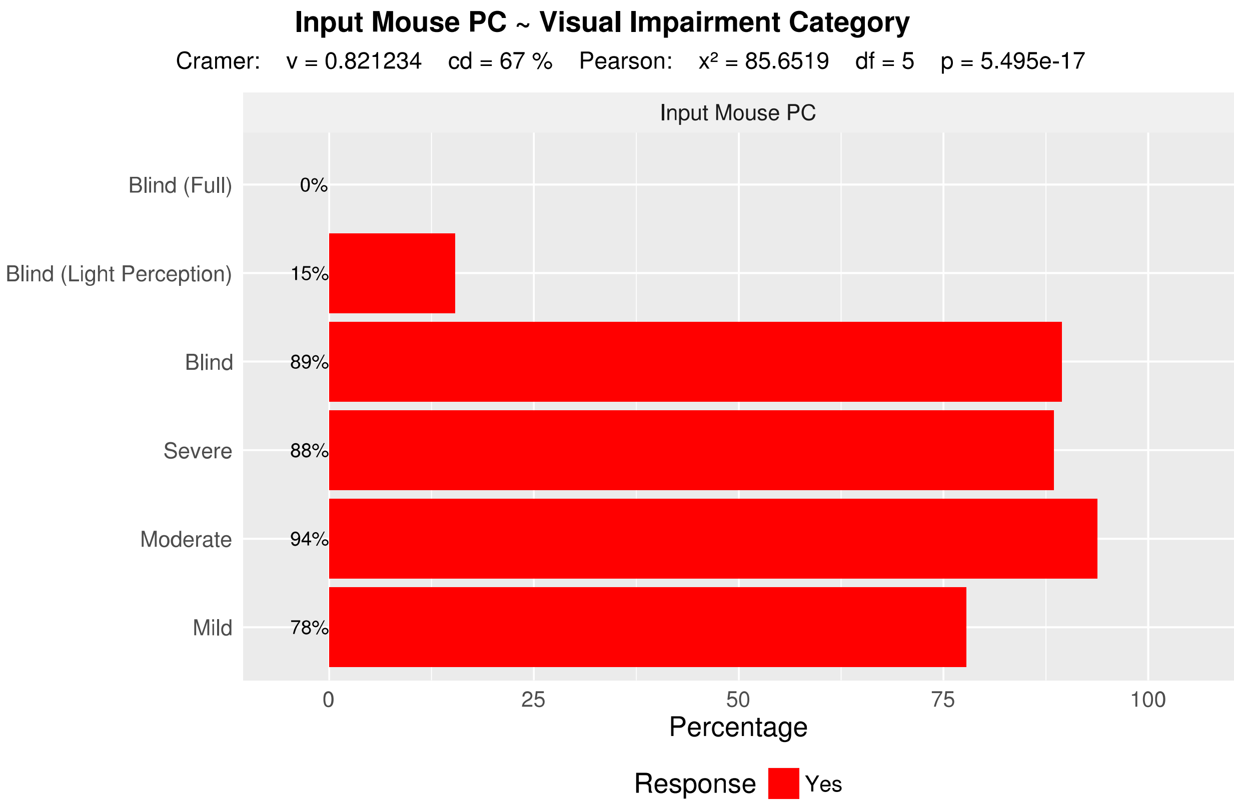


Figure 4.33: Mouse for PC in correlation to Visual Impairment Categories

78 per cent of persons with a mild visual impairment as well as 94 per cent of persons with a moderate visual impairment, 88 per cent of persons with s severe visual impairment, 89 per cent of blind persons with a visus worse than 0.05 and only 15 per cent of blind persons with light perception are using mouse for desktop computers. Mouse is not used by the category of fully blind persons without light perception because for using mouse some visual performance is required: In order to get an additional value, the person with disability must be able to seeing the current position of the mouse pointer on the screen as well as the corresponding document structure element it points at.

Gestures on touch screen or touch pad are used seldom for desktop computers at the moment by only 6 per cent of the respondents. The use of this input modality may highly increase in the future because of the emerging number of applications where gestures on touch screen or multi touch pad is used as the primary source of input.

Although there are many solutions for speech input like Linguatec Voice Pro [VoicePro] or Nuance Dragon [Dragon] available for desktop computers, speech is used very seldom by only 5 per cent of the persons concerned.

Joystick is also used very seldom for desktop computers by only 3 per cent of the persons concerned because for desktop computers there are few applications supporting the use of joystick for the navigation within structured documents.

For the reading of and the navigation within structured documents on mobile and wearable devices, as with the category of desktop computers described above, blind and visually impaired persons are using a combination of different input modalities. Figure 4.34 shows the different input modalities reported by the persons concerned as well as the percentages of how often each input modality is used:

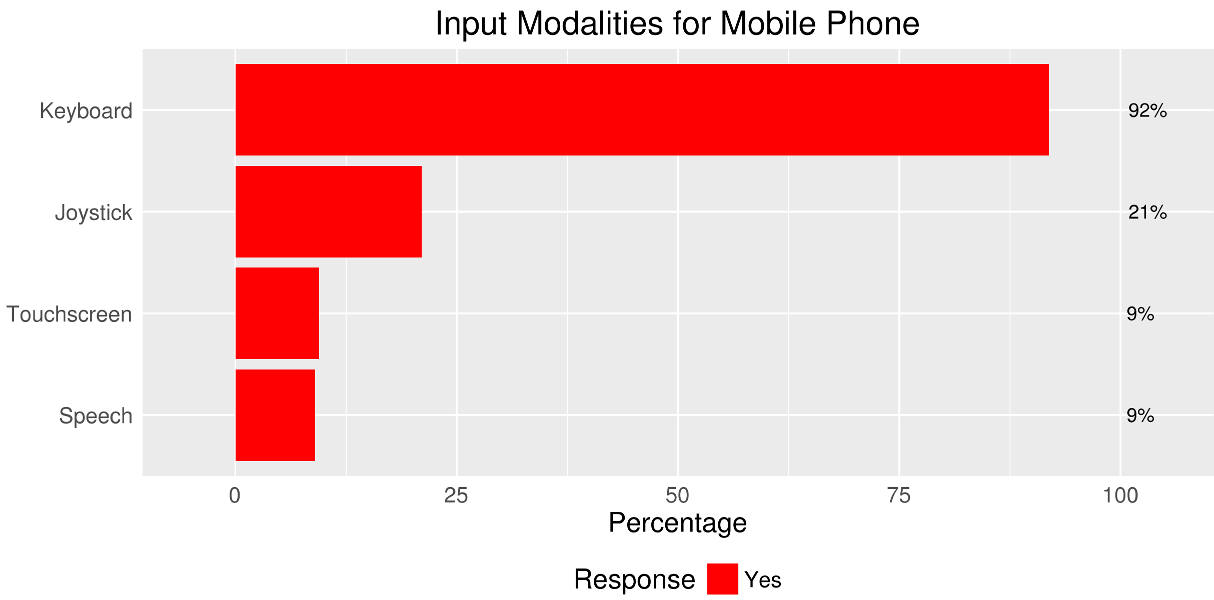


Figure 4.34: Mobile Input Modalities used for Navigation within Documents

Keyboard is used by 92 per cent of the blind and visually impaired persons for the reading of and the navigation within structured documents, because most assistive technologies like screen reader or screen magnifier provide numerous key sequences for the navigation within structured documents, for example for jumping from one heading to the next.

21 per cent of the persons concerned are using joystick for the reading of and the navigation within structured documents. On mobile and wearable devices joystick is used very often for basic navigation tasks like scrolling forward and backward.

Gestures on touch screen are at the moment used seldom by 9 per cent of blind and visually impaired persons. This input modality may highly increase in the future because of the emerging category of mobile and wearable devices like smart phones or smart tablets where the multi-touch screen is the primary source of input.

The use of touch screen on mobile and wearable devices depends to 76 per cent on the screen reader product used, as shown by a very strong bivariate correlation with $v = 0.87$ and a very high level of statistical significance with $p = 3.21e-9$. Figure 4.35 shows the use of the touch screen input modality on mobile and wearable devices in correlation to the screen reader products used on mobile and wearable devices as well as the percentages of how often touch screen is used:

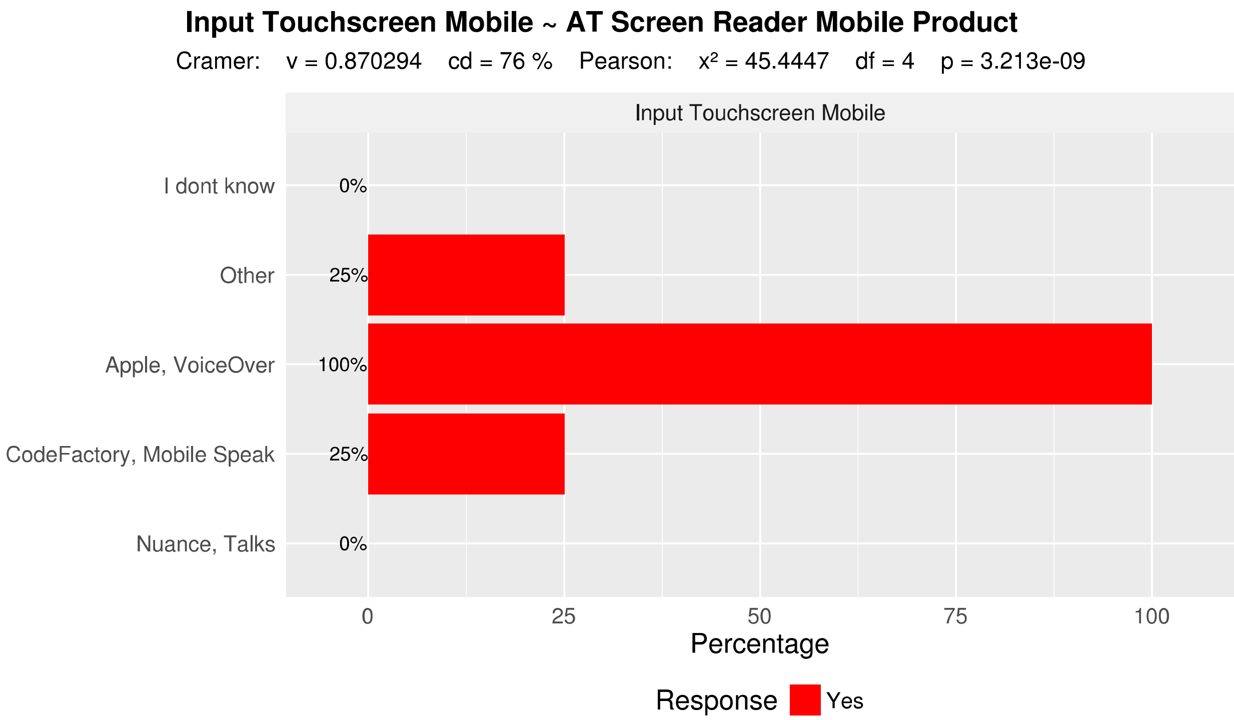


Figure 4.35: Mobile Touch Screen in correlation to Mobile Screen Reader Products

100 per cent of the persons who are using the Apple VoiceOver as well as 25 per cent of the persons using CodeFactory Mobile Speak and another 25 per cent of the persons using another screen reader product are using touch screen because these screen reader products provide gestures for the reading of an the navigation within structured documents on mobile and wearable devices. The category of persons using Nuance Talks and persons who do not know which screen reader product they are using do not use touch screen at all because these screen reader products do not provide gestures.

Another moderate bivariate correlation with v = 0.46 and a very high statistical significance level of p = 1.73e-9 shows that the use of touch screen on mobile and wearable devices is to 21 per cent dependent on the use of touch screen for desktop computers. Figure 4.36 shows the use of touch screen on mobile and wearable devices in correlation to the use of touch screen for desktop computers as well as the percentages of how often touch screen in used:

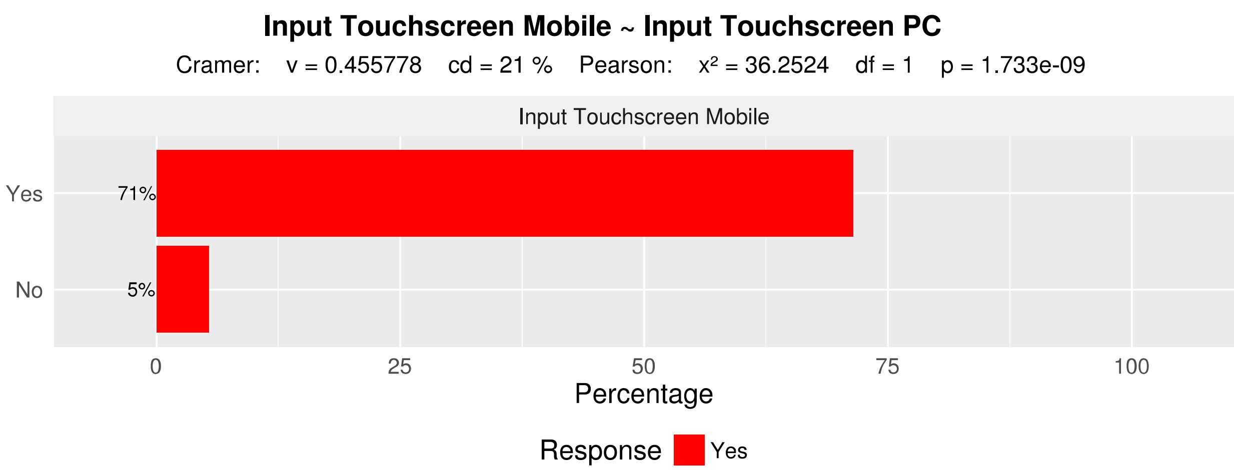


Figure 4.36: Mobile Touch Screen in correlation to Touch Screen for PC

71 per cent of the persons who are using touch screen for desktop computers and only 5 per cent of the persons not using touch screen for desktop computers also use touch screen on mobile and wearable devices.

Although there are many solutions for speech input like Linguatec Voice Pro [VoicePro] or Nuance Dragon [Dragon], available for both desktop computers as well as for mobile and wearable devices, speech is used very seldom by another 9 per cent of the persons concerned.

In order to use assistive technologies like screen reader or screen magnifier efficiently, many key sequences have to be known by heart and these key sequences differ between the different products available on the market. An exception is Apple VoiceOver [VoiceOver]. Novel input modalities like gestures on multi-touch screen or multi-touch pad, motion or speech, provided by the emerging category of mobile and wearable devices out of the box, should be employed in addition to the traditional input modalities like keyboard and mouse in a novel concept for an assistive technology in support of blind and visually impaired persons for the reading of structured documents because the persons concerned would wish to have a more easy, intuitive and standardised way for the reading of and the navigation within structured documents.

4.3.6 Structure Elements

The logical document structure in general is a very important part of a structured document. However, some structure elements are more important for blind and visually impaired persons as to the reading of and the navigation within structured documents. than others. Figure 4.37 shows the logical structure elements reported by the persons concerned, which they use for the reading of and the navigation within structured documents, as well as the percentages of how important each structure element is for them:

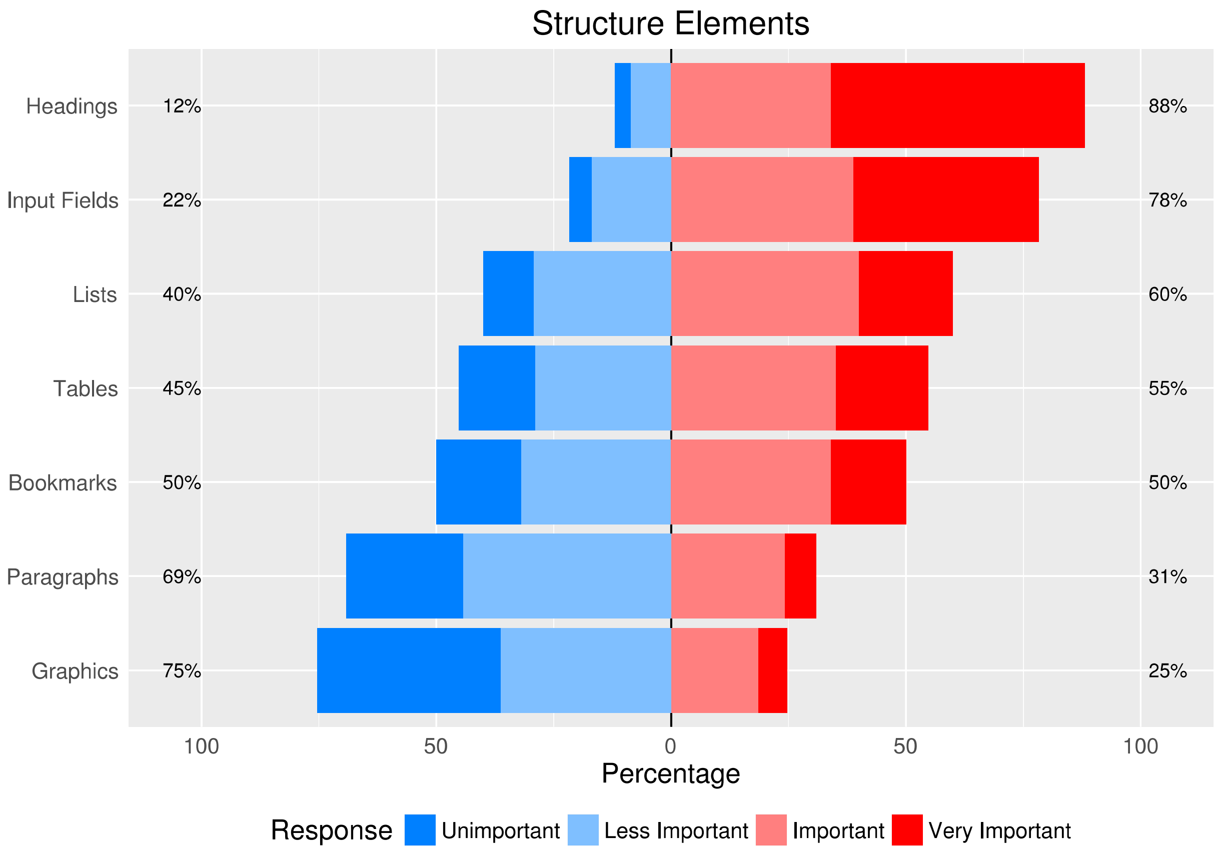


Figure 4.37: Structure Elements used by Research Subjects for Navigation

The most important structure element for blind and visually impaired persons are headings because in most cases they are used as the primary element for structuring text. For 88 per cent of them headings are important and very important and for only 12 per cent of them headings are less important and unimportant.

The importance of headings depends to 12 per cent on the use of the HTML digital content format, as shown by a weak bivariate correlation with v = 0.34 and a vary high level of statistical significance with p = 5.19e-8. Figure 4.38 show the importance of the heading structure element in correlation to the use of the HTML digital content format as well as the percentages of how important headings are:

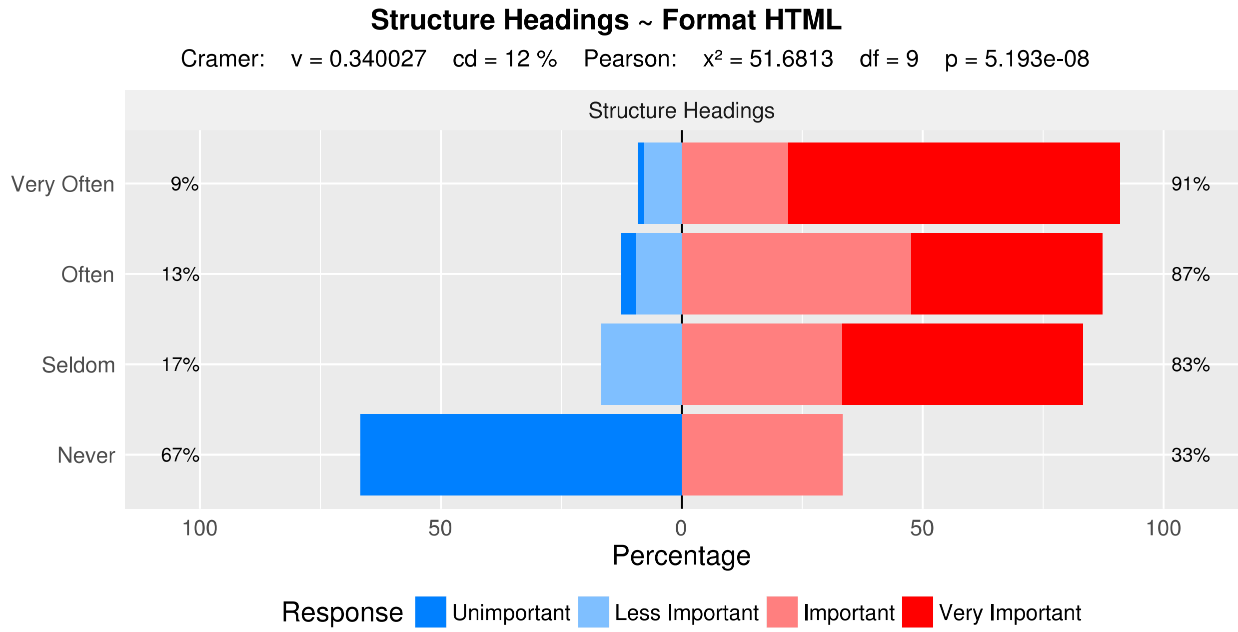


Figure 4.38: Importance of Headings in correlation to HTML Digital Format

91 percent of persons receiving structured documents in the HTML format very often find headings important and very important and for 9 per cent of them headings are less important and unimportant. 87 per cent of persons who receive structured documents in the HTML format often find headings important and very important and for only 13 per cent of them headings are less important and unimportant. 83 per cent of persons receiving structured documents in the HTML format seldom find headings important and very important and only for only 17 per cent of them headings are less important and unimportant. Finally, only 33 per cent of the persons who never receive structured documents in the HTML format find headings important and for 67 per cent of them headings are unimportant. This correlation arises due to the fact that headings are used very extensively in highly structured HTML documents and websites.

The importance of headings is followed by input fields which are the second most important structure element, assumed that the structured document in question is a form where the user has to input some information. 78 per cent of the persons concerned reported input fields to be important and very important and only 22 per cent of them find input fields less important and unimportant.

The importance of input fields is to 18 per cent dependent on the use of the mouse input modality for desktop computers, as shown by a moderate bivariate correlation with v = 0.42 and a very high level of statistical significance with p = 2.02e-5. Figure 4.39 shows the importance of the input fields structure element in correlation to the use of the mouse input modality for desktop computers as well as the percentages of how important input fields are:

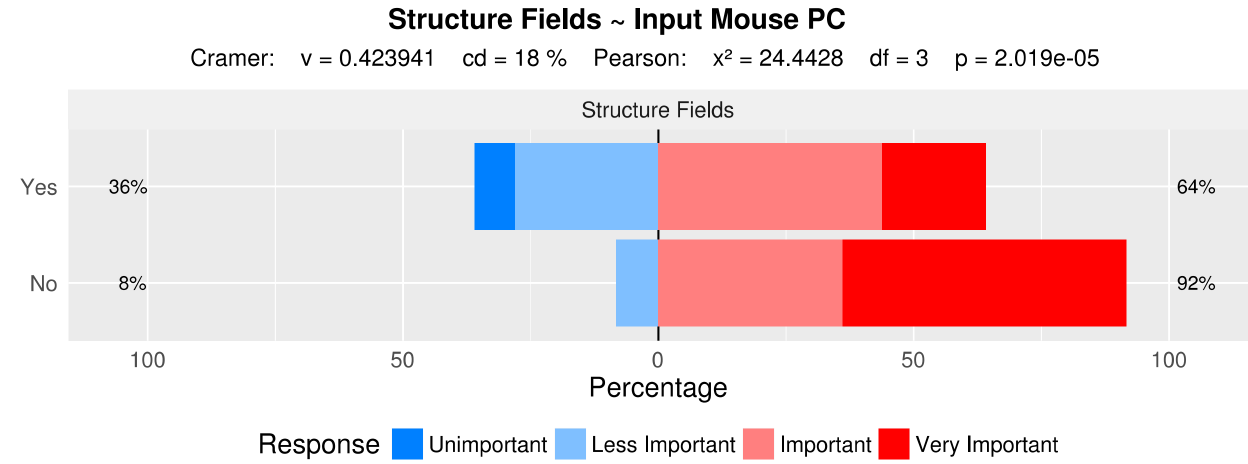


Figure 4.39: Importance of Input Fields in correlation to Mouse Input Modality for PC

92 per cent of the respondents who are unable to using mouse for desktop computers find input fields important and very important and for only 8 per cent of them input fields are less important and unimportant. 64 per cent of the persons who are able to use mouse for desktop computers find input fields important and very important and for 36 per cent of them input fields are less important and unimportant. This correlation arises due to the fact that persons who are unable to using mouse are reliant that input fields are semantically contained within the logical document structure in order to be able to focus them while persons who are using mouse can simply click on an input field even when the input field is not semantically contained within the logical document structure.

The importance of input fields is followed by the importance of lists. 60 per cent of the respondents find lists important and very important for the reading of and the navigation within structured documents and only 40 per cent of them find lists less important and unimportant.

The importance of lists is followed by the importance of tables. 55 per cent of the blind and visually impaired persons find tables important and very important and 45 per cent of them reported tables to be less important and unimportant.

The importance of tables depends to 23 per cent on the importance of lists, as shown by a moderate bivariate correlation with v = 0.48 and a very high level of statistical significance with p = 3.83e-18. Figure 4.40 shows the importance of the table structure element in correlation to the list structure element as well as the percentage of how important tables are:

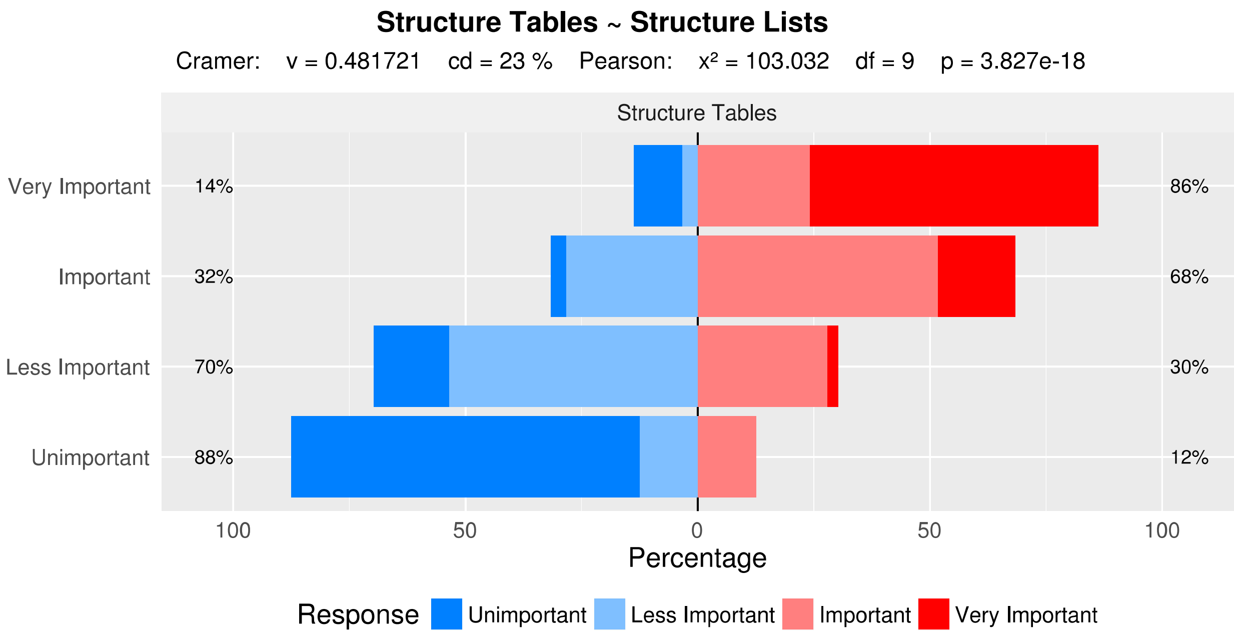


Figure 4.40: Importance of Tables in correlation to Importance of Lists

86 per cent of the respondents who find lists important and very important also find tables important and very important and for only 14 per cent of them tables are less important and unimportant. 68 per cent of the persons who find lists important also find tables important and very important and for only 32 per cent of them tables are less important and unimportant. On the other hand 70 per cent of the respondents who find lists less important also find tables less important or unimportant and 30 per cent of them find tables important or very important. Finally, 88 per cent of the respondents who find lists unimportant also find tables less important and unimportant and for only 12 per cent of them tables are important and very important.

The importance of tables is followed by the importance of bookmarks. 50 per cent of the persons concerned find bookmarks important and very important and for another 50 per cent of them bookmarks are less important or unimportant because in many cases they are used to convey the same information like headings.

The importance of bookmarks is to 5 per cent dependent on the use of the PDF digital content format, as shown by a weak bivariate correlation with $v = 0.21$ and high level of statistical significance with $p = 0.017$. Figure 4.41 shows the importance of the bookmark structure element in correlation to the use of the PDF digital content format as well as the percentages of how important bookmarks are:

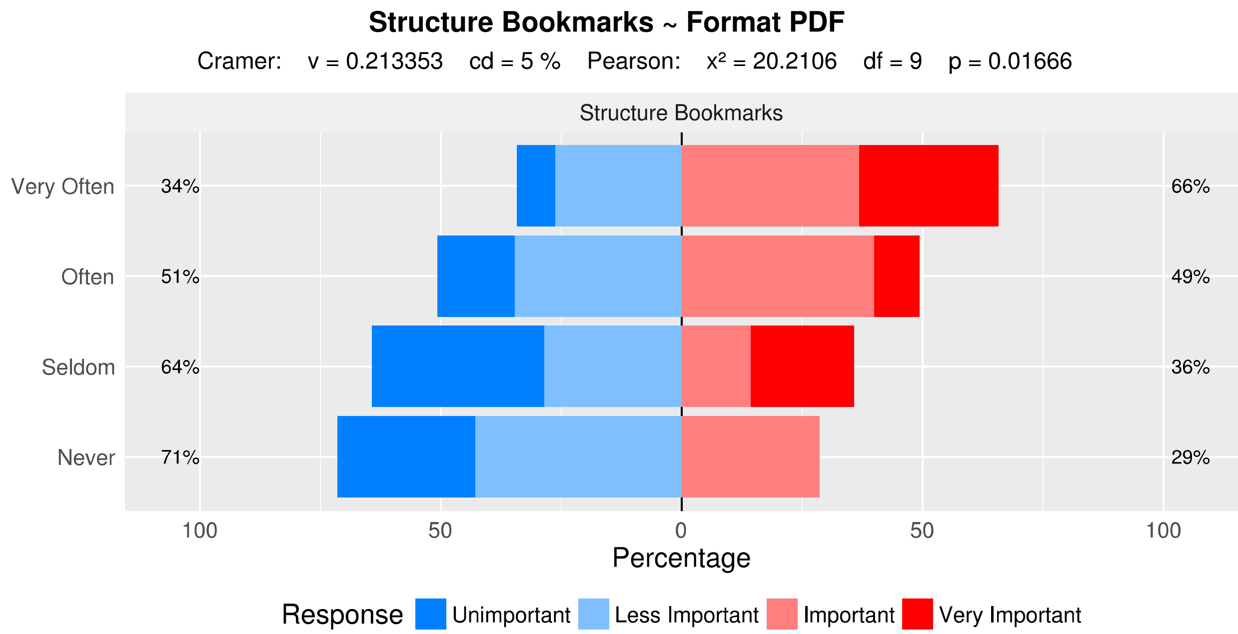


Figure 4.41: Importance of Bookmarks in correlation to PDF Digital Format

66 per cent of the respondents who receive structured documents very often in the digital content format PDF find bookmarks important and very important and for only 34 per cent of them bookmarks are less important or unimportant. 49 per cent of the respondents who receive structured documents often in the PDF format find bookmarks important and very important and for 51 per cent of them bookmarks are less important or unimportant. On the other hand 64 per cent of the respondents who receive PDF documents seldom find bookmarks less important or unimportant and for 36 per cent of them bookmarks are important or very important. And finally 71 per cent of the respondents who never receiver documents in the PDF format also find bookmarks less important or unimportant and for only 29 per cent of them bookmarks are important and very important. This correlation arises due to the fact that bookmarks are typically used very extensively in structured PDF documents.

The importance of bookmarks is followed by the importance of paragraphs. Paragraphs are regarded as less important because apart from the length of the text contained in them, they do not provide an additional information about the structure of a document. Only 31 per cent of the respondents find paragraphs important or very important and for 69 per cent of them paragraphs are less important or unimportant.

The least important element for the blind and visually impaired persons are graphics because in many cases they are not able to seeing them visually and therefore they do not provide any value for them apart from a text alternative, which may be provided by the author. 75 per cent of them find graphics less important and unimportant and for only 25 per cent of them graphics are important or very important.

The importance of graphics depends to 14 per cent on the use of the mouse input modality for desktop computers, as shown by a moderate bivariate correlation with v = 0.37 and a very high level of statistical significance with p = 3.704e-4. Figure 4.42 shows the importance of the graphic structure element in correlation to the use of the mouse input modality for desktop computers as well as the percentages of how important graphics are:

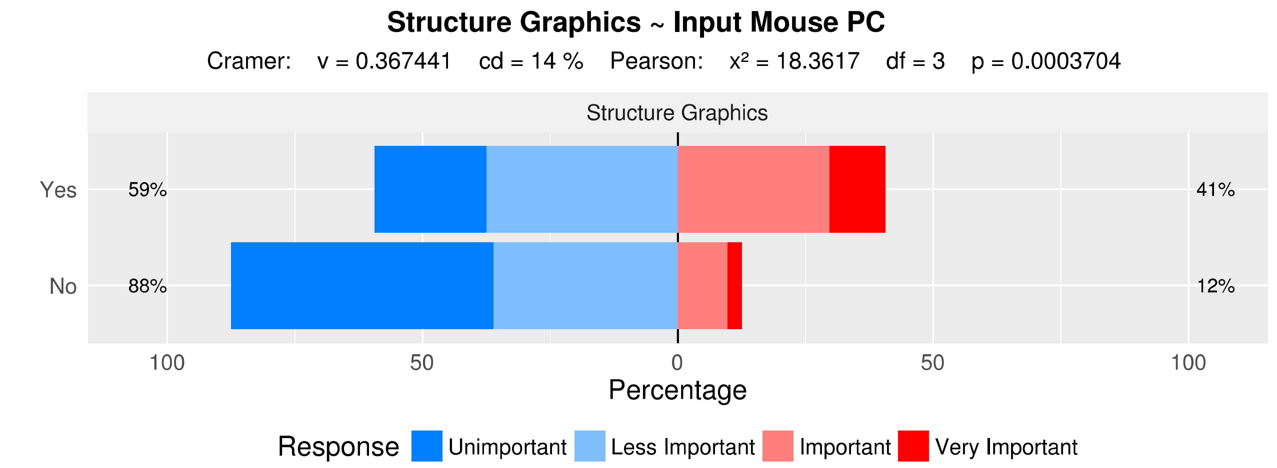


Figure 4.42: Importance of Graphics in correlation to Mouse Input Modality for PC

41 per cent of the persons who are using mouse for desktop computers find graphics important and very important and for 59 per cent of them graphics are less important or unimportant. On the other hand only 88 per cent of the persons who are unable to using mouse for desktop computers also find graphics less important or unimportant and for only 12 per cent of them graphics are important and very important. This correlation arises due to the fact that for persons who are unable to using mouse, graphics do not provide an addition value apart from a text alternative which may be provided by the author.

The logical document structure plays a very important role as to the reading of and the navigation within structured documents. A novel concept should be developed for non-visual presentation of the logical structure elements employing novel output modalities like auditory icons or vibration feedback, provided by the emerging category of mobile and wearable devices out of the box, in addition to the traditional output modalities like speech synthesis or Braille displays used by state of the art screen readers because all of the participants would wish to have a better overview over the logical document structure of structured documents.

4.3.7 Accessibility Problems

Most structured documents available on the Internet are not or only partially accessible for blind and visually impaired persons because all of the persons concerned reported that they receive documents which they cannot access. Figure 4.43 shows the different accessibility problems reported by the persons concerned as well as the percentages of how often each accessibility problem occur to them. These different accessibility problems are briefly described below.

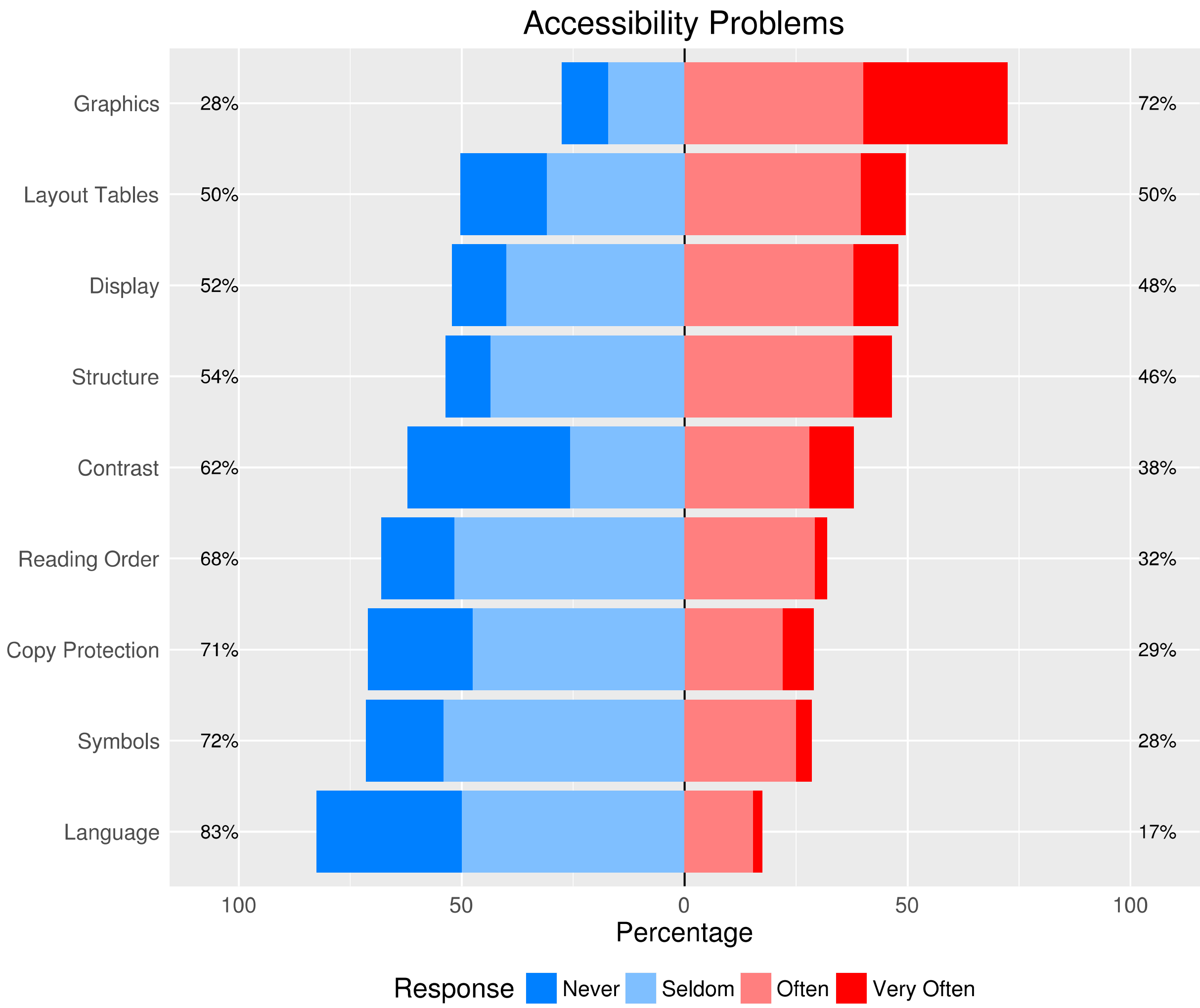


Figure 4.43: Accessibility Problems of Structured Documents for Research Subjects

Graphics which do not provide a text alternative is the problem occurring to blind and visually impaired persons most often. In order for a graphical content to provide value on a non-visual medium like speech synthesis or Braille display, an alternative text or long description must be provided. To 72 per cent of the persons concerned graphics without text alternative occur often and very often and to only 28 per cent of them this problem occurs seldom or never.

The occurrence of this graphics problem depends to 13 per cent on the use of screen readers for desktop computers, as shown by a moderate bivariate correlation with v = 0.36 and a very high level of statistical significance with p = 4.83e-4. Figure 4.44 shows the occurrence of the graphics accessibility problem in correlation to the use of the screen reader assistive technology for desktop computers as well as how often this problem occurs:

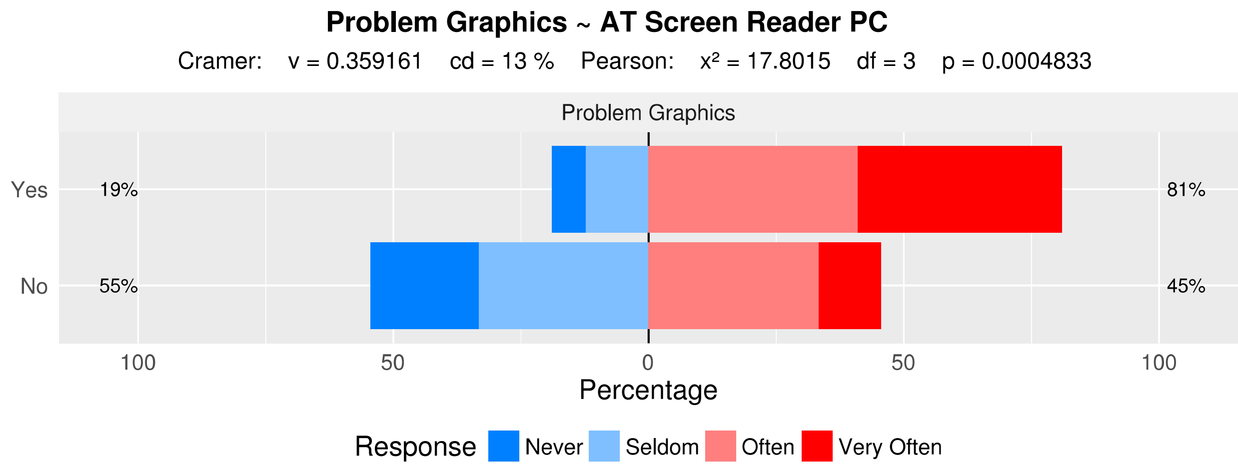


Figure 4.44: Graphics Problem in correlation to Screen Readers for PC

To 81 per cent of the respondents who are using a screen reader for desktop computers the graphics problem happens often and very often and to only 19 per cent of them this problem happens seldom or never. On the other hand to 55 per cent of the persons who are not using a screen reader for desktop computers this problem happens seldom or never and to only 45 per cent of them this problem happens often and very often. This correlation arises due to the fact that persons not using screen readers and who are able to seeing graphical content visually, do not require a text alternative.

Tables should only be used for the storage of semantically corresponding data and not for layout purposes, what is called as layout tables. In order to enable non-visual navigation within a table, heading rows and heading columns must be defined and semantically contained within the logical document structure. Problems with tables occur to 50 per cent of the blind and visually impaired persons often and very often and to another 50 per cent of them, this problem occurs seldom or never.

The occurrence of this table problem depends to 10 per cent on the use of screen readers for desktop computers, as shown by a weak bivariate correlation with v = 0.31 and a very high level of statistical significance with p = 4.69e-3. Figure 4.45 shows the occurrence of the table accessibility problem in correlation to the use of the screen reader assistive technology for desktop computers as well as how often this problem occurs:

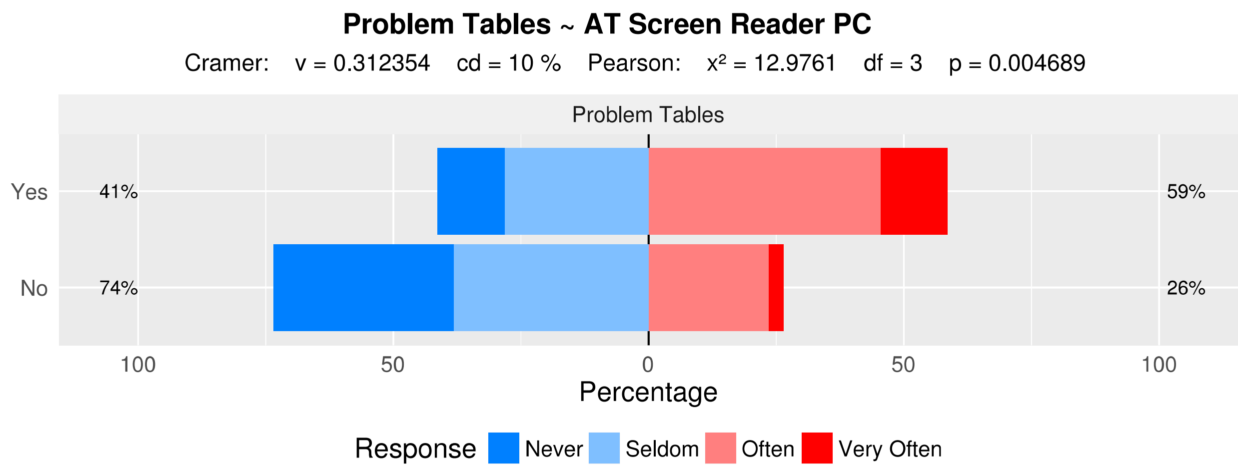


Figure 4.45: Table Problem in correlation to Screen Readers for PC

To 59 per cent of the respondents who are using a screen reader for desktop computers the table problem happens often and very often and to only 41 per cent of them this problem happens seldom or never. On the other hand to 74 per cent of the persons who are not using a screen reader for desktop computers this problem happens seldom or never and to only 26 per cent of them this problem happens often and very often. This correlation arises due to the fact that for persons not using screen readers and who are able to reading tables visually, the logical semantical table structure is not required and layout tables are invisible for them anyway.

As with all structure elements in general, display accentuations within the document's content must be semantically contained within the logical document structure in addition to their visual physical presentation (for example using bold font or a different color) to be accessible. To 48 per cent of the blind and visually impaired persons this problem occurs often and very often and to 52 per cent of them this problem occurs seldom or never.

In order for the structure of a document to be accessed and presented on a non-visual medium like speech synthesis or Braille display, the logical document structure must be semantically contained within the document in addition to the visual physical structure (for example using a bigger font for a heading). 46 per cent of the persons concerned reported to receive documents with inaccessible structure often and very often and 54 per cent of them receive such documents seldom or never.

The occurrence of this structure problem depends to 16 per cent on the occurrence of the display accentuations problem, as shown by a moderate bivariate correlation with v = 0.4 and a very high level of statistical significance with p = 1.22e-10. Figure 4.46 shows the occurrence of the structure accessibility problem in correlation to the display accentuations accessibility problem as well as how often this problem occurs:

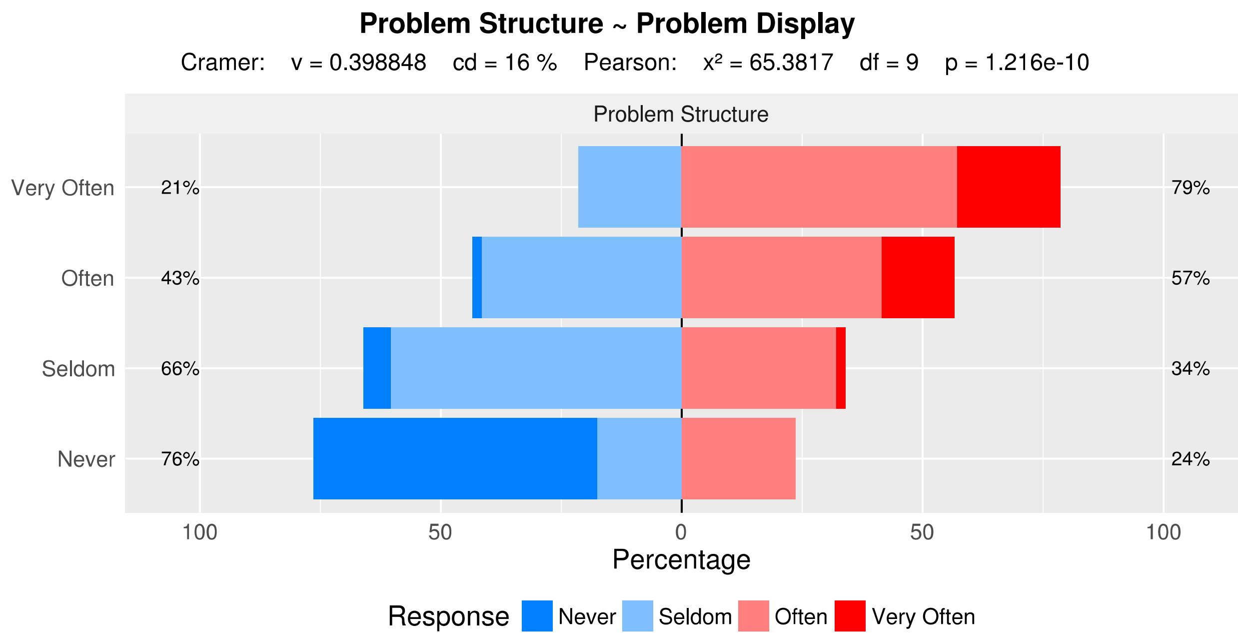


Figure 4.46: Structure Problem in correlation to the Display Problem

To 79 per cent of respondents having problems with display accentuations very often the structure problem also happens often and very often and to only 21 per cent of them this problem happens seldom. To 57 per cent of respondents having problems with display accentuations often the structure problem also happens often and very often and to 43 per cent of them this problem happens seldom or never. On the other hand to 66 per cent of the persons having problems with display accentuations seldom or never the structure problem happens seldom or never and to only 34 per cent of them this problem happens often and very often. Finally, to 76 per cent of persons who never have problems with display accentuations the structure problem happens seldom or never and to only 24 per cent of them this problem happens often. This correlation arises due to the fact that if the logical document structure is not semantically contained within a document, display accentuations are not semantically contained too because they are port of the logical document structure.

If the contrast ratio between the foreground and the background color is not sufficient a content element (text or graphic) can visually not be read correctly. 38 per cent of the respondents are affected by the problem of insufficient contrast often and very often and 62 per cent of them are affected by this problem seldom or never.

The occurrence of this contrast problem is to 18 per cent dependent on the use of screen magnifiers for desktop computers, as shown by a moderate bivariate correlation with v = 0.43 and a very high level of statistical significance with p = 3.43e-5. Figure 4.47 shows the occurrence of the contrast accessibility problem in correlation to the use of the screen magnifier assistive technology for desktop computers as well as the percentages of how often the contrast problem occurs:

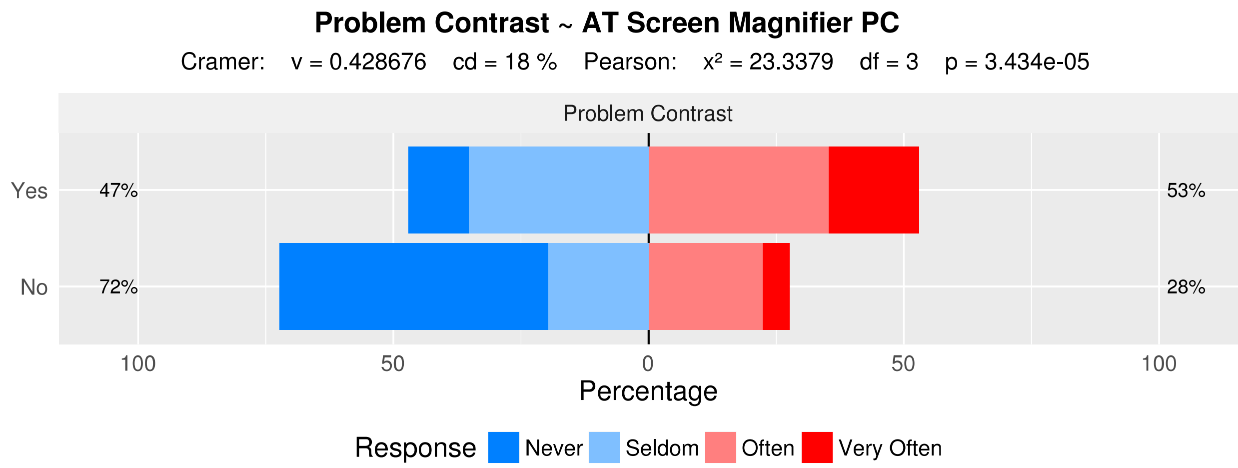


Figure 4.47: Contrast Problem in correlation to Screen Magnifiers for PC

To 53 per cent of the respondents who are using a screen magnifier for desktop computers the contrast problem happens often and very often and to 47 per cent of them this problem happens seldom or never. On the other hand to 72 per cent of the persons who are not using a screen magnifier for desktop computers this problem happens seldom or never and to only 28 per cent of them this problem happens often and very often. This correlation arises due to the fact that persons not using screen magnifiers because they do not use the visual physical document structure or because they are able to read a document visually without using any assistive technology at all, contrast is not required.

An important requirement for non-visual navigation within a structured document is that the reading order is semantically defined correctly. If the reading order is incorrect, the document's content does not make sense for the reader. The problem of incorrect reading order occurs to 32 per cent of the blind and visually impaired persons often and very often and to 68 per cent of them this problem occurs seldom or never.

The occurrence of this reading order problem is to 16 per cent dependent on the use of mouse for desktop computers, as shown by a moderate bivariate correlation with v = 0.40 and a very high level of statistical significance with p = 7.07e-5. Figure 4.48 shows the occurrence of the reading order accessibility problem in correlation to the use of the mouse input modality for desktop computers as well as the percentages of how often the reading order problem occurs:

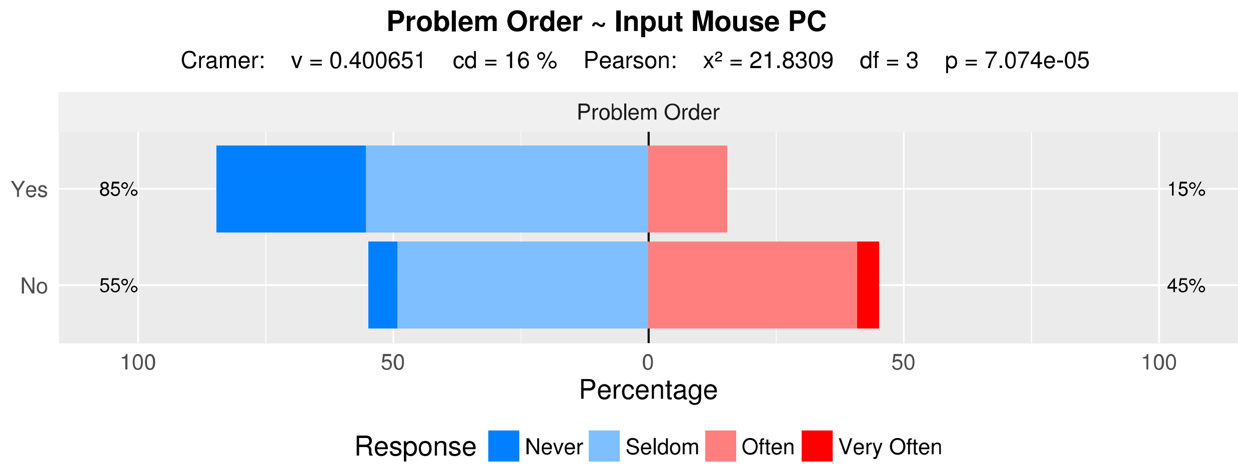


Figure 4.48: Reading Order Problem in correlation to the Mouse Input Modality for PC

To 45 per cent of the respondents who are unable to using mouse for desktop computers the problem of wrong reading order happens often and very often and to 55 per cent of them this problem happens seldom or never. On the other hand to 85 per cent of the persons who are using mouse for desktop computers this problem happens seldom or never and to only 15 per cent of them this problem happens often and very often. This correlation arises due to the fact that persons who are unable to using mouse are reliant on a semantically correct reading order to be able to navigate within the logical document structure using alternative non-visual input modalities like keyboard.

If the user agent does not grant access to the documents structure and content, the assistive technology is unable to present this information to the reader with a disability. Especially with the copy protection of recent versions of E-Book reading software this has become an issue. 29 per cent of the persons concerned are affected by the problem of copy protection often and very often and 71 per cent of them are affected by this problem seldom or never.

for the categories of severe visually impaired and blind persons. Mild and moderate visually impaired persons on the other hand are less affected because they may see the visual presentation without the use of an extra assistive technologies.

The occurrence of this copy protection problem is to 8 per cent dependent on the use of screen readers for desktop computers, as shown by a weak bivariate correlation with v = 0.28 and a very high level of statistical significance with p = 0.015. Figure 4.49 shows the occurrence of the copy protection accessibility problem in correlation to the use of the screen reader assistive technology for desktop computers as well as the percentages of how often the copy protection problem occurs:

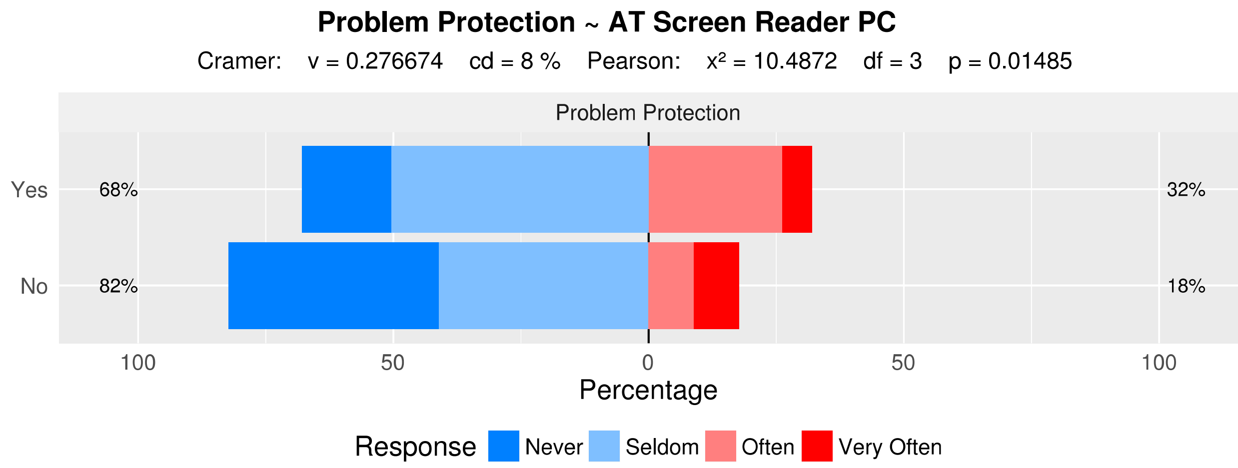


Figure 4.49: Copy Protection Problem in correlation to Screen Readers for PC

To 32 per cent of the respondents who are using a screen reader for desktop computers the copy protection problem happens often and very often and to 68 per cent of them this problem happens seldom or never. On the other hand to 82 per cent of the persons who are not using a screen reader for desktop computers the copy protection problem happens seldom or never and to only 18 per cent of them this problem happens often and very often. This correlation arises due to the fact that persons not using screen readers because they are able to reading the structured document visually, do not require access to the logical document structure semantically.

Some special character symbols cannot be displayed correctly by non-visual output modalities like speech synthesis or Braille displays or in enlarged visual form. To 28 per cent of the respondents the problem of unrecognizable symbols occurs often and very often and to 72 per cent of them this problem occurs seldom or never.

In order to present a text using speech, the language of the text must be known by the assistive technology for selecting the correct speech synthesis. If a text is spoken using a speech synthesis of a wrong language, because the language of the text is not or incorrect defined within the logical document structure, the text cannot be understood by the reader. The language can be set globally for the whole document or individual for each structural element containing text. The problem of incorrect language happens often and very often to 17 per cent of the blind and visually impaired persons and to 83 per cent of them this problem happens seldom or never.

At the moment most structured documents available on the Internet are not or not fully accessible for blind and visually impaired persons because all of the persons concerned reported that they receive documents which they cannot access. This is not a problem of missing scientific concepts more than that these concepts did not move into the praxis. To improve the current situation authors of structured documents available on the Internet like web developer, web designer and web redactors should be highly encouraged to follow certain accessibility standards like the W3C Web Content Accessibility Guidelines WCAG 2.1 (Kirkpatrick, O'Connor, Campbell & Cooper, 2015). However, most authors do not have the knowledge on how to produce accessible documents. Therefore tools like plug-ins for authoring tools should be developed for checking the accessibility of a document and helping the author to fixing certain accessibility issues (Darvishy, 2018).

4.3.8 Accessibility Solutions

If the blind and visually impaired persons are not able to read a structured document provided to them because of one or more of the accessibility problems described in the previous Chapter 4.3.7, they take on different accessibility solutions. Figure 4.50 shows the different accessibility solutions reported by the persons concerned as well as the frequencies and percentages of respondents which take on each accessibility solution. These accessibility solutions are briefly described below.

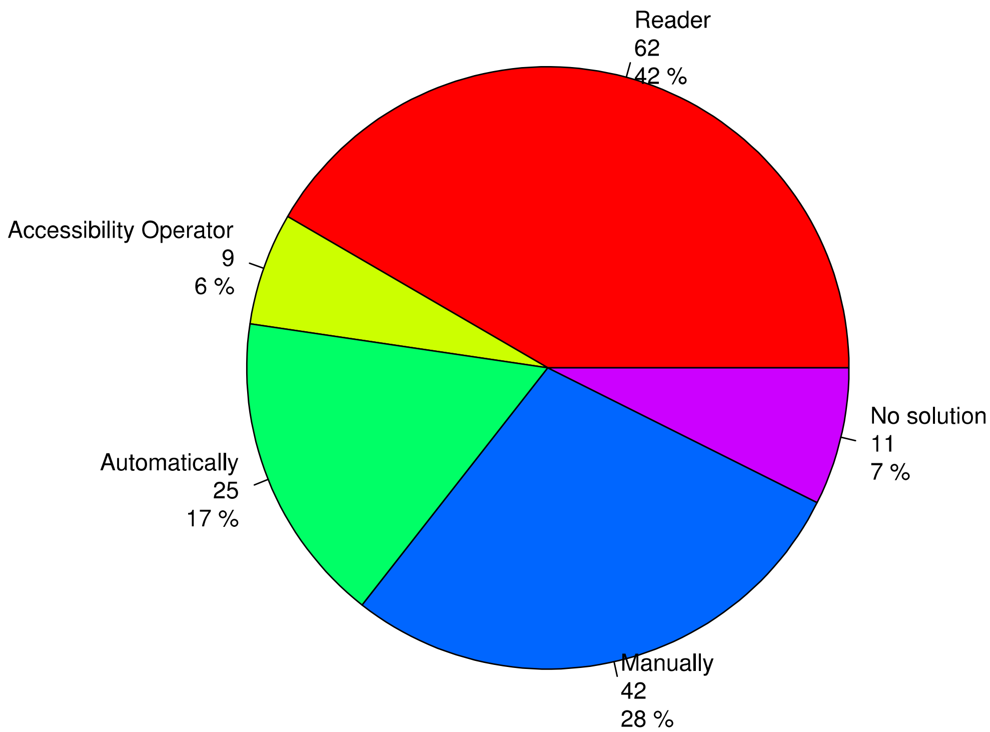


Figure 4.50: Accessibility Solutions Blind and Visually Impaired Persons are taking on

Most blind and visually impaired persons, namely 42 per cent of them, hand the inaccessible document to another person without disability, the so called reader, who reads the document to them because in many cases this is the most easiest, quickest and efficient solution. The advantages of the reader solution are that there is neither technical equipment nor computer affinity required by the person with disability. The disadvantages of this solution are that a reader person must be available at a time, otherwise the document will not be immediately available to the person with disability.

28 per cent of the persons concerned make the document manually accessible by themselves with high effort. The advantages of this manual solution is that the document is immediately available for the person with disability. The disadvantage of this solution is that there is some computer affinity of the person with disability required. In addition some visual performance may also be required.

A hardware device or software, which aims to make the document accessible automatically for example by performing object character recognition (OCR), is used by 17 per cent of the respondents. The commercially available products Nuance OmniPage [OmniPage], ABBY FineReader [FineReader] and Freedom Scientific OpenBook [OpenBook] are very popular for that purpose. The advantage of the automated solution is, as with the manual solution described above, that the resulting automatically processed document is immediately available for the person with disability. The disadvantage of this solution is that the automatic document recognition may be defective and therefore the quality of the resulting automatically processed document may not be as high as with a manually processed version by an accessibility operator.

6 per cent of the blind and visually impaired persons hand the inaccessible document to another person without disability, the so called accessibility operator, who makes the document manually accessible for them by converting graphics into text, defining or correcting the logical document structure semantically, providing text alternatives for graphics, etc. The advantage of the accessibility operator solution is that the accessibility of the resulting manually processed document is of a very high quality. The disadvantage of this solution is, as with the reader solution described above, that the document is not available immediately for the person with disability. The person with disability has to wait until the document has been manually processed by the accessibility operator.

Finally, 7 per cent of the persons concerned reported not to having a solution at all. If the receive an inaccessible document, they simply do not read the document. Instead they return it to the author with the notice that they are unable to reading the document and asking the author to provide the document to them in a more accessible form.

The use of these accessibility solutions depends to 5 per cent on the computer affinity a specific person comes with, as shown by a weak bivariate correlation with v = 0.23 and a high level of statistical significance with p = 0.049. Figure 4.51 shows the use of the different accessibility solutions in correlation to the computer affinity a specific person comes with as well as the frequencies and percentages of respondents which take on each accessibility solution:

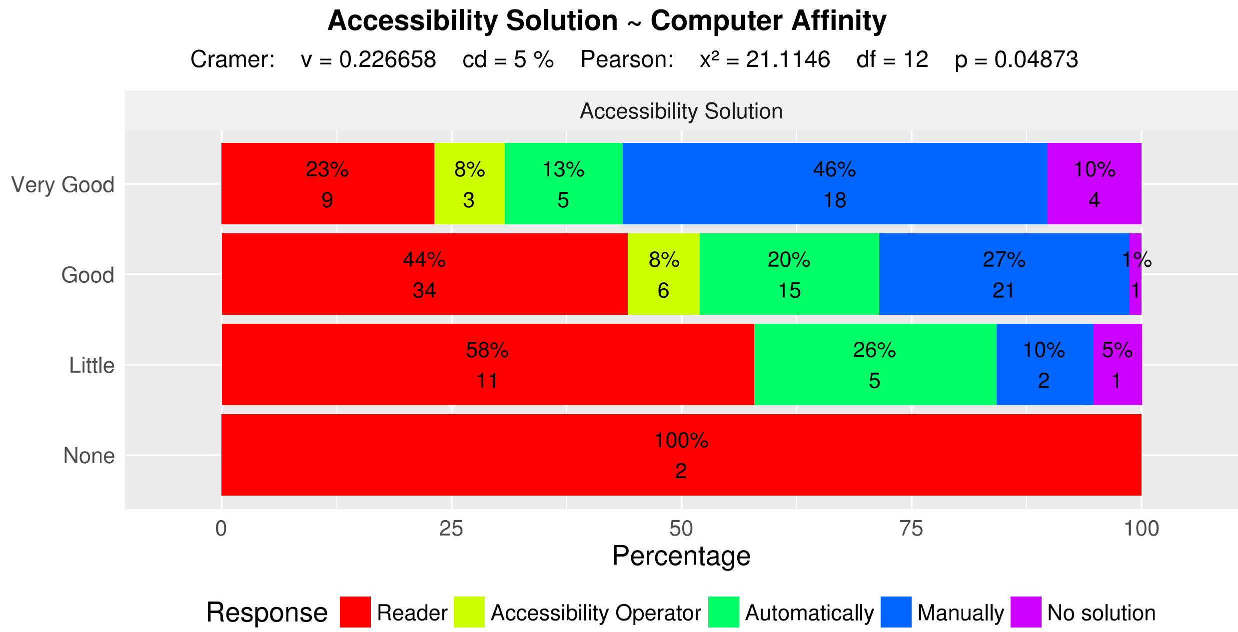


Figure 4.51: Accessibility Solutions in correlation to the Computer Affinity

23 per cent of persons with very good computer affinity as well as 44 per cent of persons with good computer affinity, 58 per cent of persons with little computer affinity and 100 per cent of persons who are not computer affine at all take on the reader solution. The manual solution is taken on by 13 per cent of persons with very good computer affinity, 20 per cent of persons with good computer affinity and 26 per cent of persons with little computer affinity. 46 per cent of persons with very good computer affinity, 27 per cent of persons with good computer affinity and only 10 per cent of persons with little computer affinity take on the automatic solution. The accessibility operator solution is taken on by only 8 per cent of persons with very good computer affinity as well as by another 8 per

cent of persons with good computer affinity. Finally, only 10 per cent of persons with very good computer affinity, 1 per cent of persons with good computer affinity and 5 per cent of persons with little computer affinity have no solution. This correlation is due to the fact that persons who are less computer affine are unable to making a document manually accessible by themselves or using a hardware device or software, which aims to make the document accessible automatically.

4.3.9 Hardware Devices

At the moment, the blind and visually impaired participants are using a variety of different hardware devices for the reading of structured documents and they are planning to purchase even more of them in the future within the next 2 to 3 years. Figure 4.52 shows the different hardware devices reported by the persons concerned, which they use for the reading of structured documents at the moment and Figure 4.53 in the future within the next 2 to 3 years, as well as the percentages of how often each hardware device is used by them:

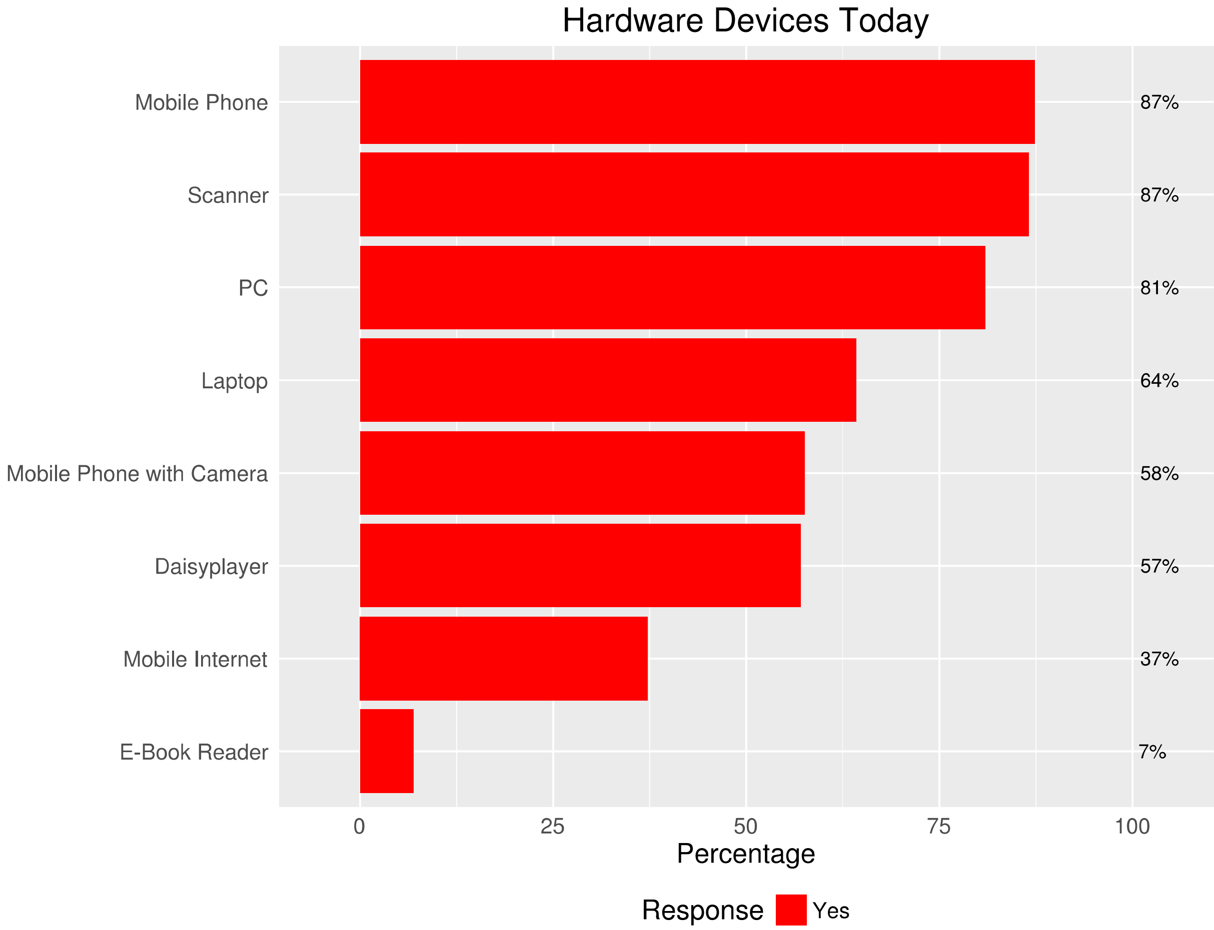


Figure 4.52: Hardware Devices used by Research Subjects Today

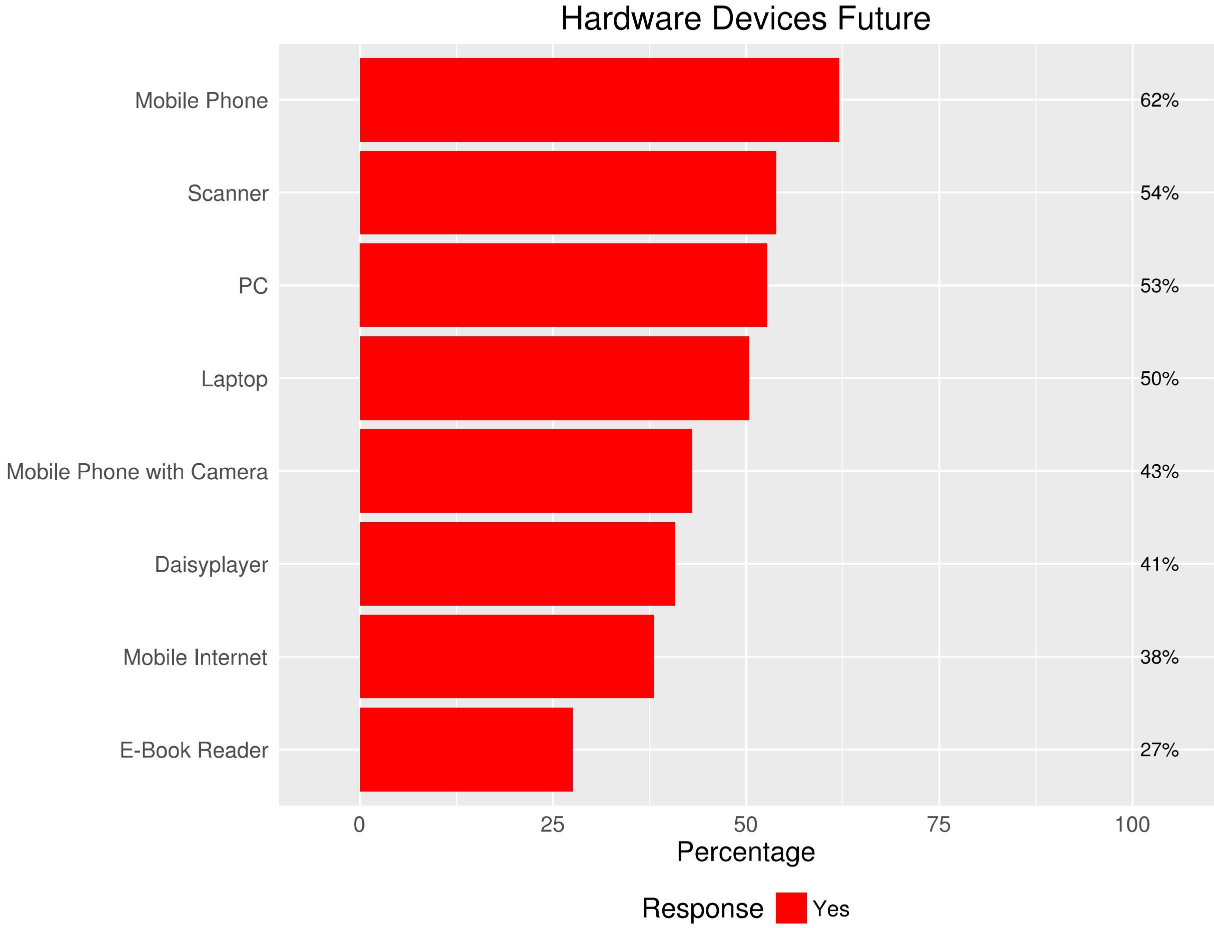


Figure 4.53: Hardware Devices used by Research Subjects in the Future

The emerging category of smart phones are used by 82 per cent of the blind and visually impaired persons and 62 per cent of them have plans to purchase a new or additional smart phone in the future within the next 2 to 3 years. Smart phones are the hardware devices most used by them because blind and visually impaired persons would wish to have their assistive technology for the reading of structured document mobile more than bound to a fixed place.

A scanner is used by 87 per cent of the persons concerned and 54 per cent of them are planning to purchase a new or additional scanner in the future within the next 2 to 3 years. They use the scanner for the processing of printed documents in order to bring them in an accessible form by using a document recognition software.

81 per cent of the respondents are using a desktop computer at home or at the workplace and 53 per cent of them are planning to purchase a new or additional desktop computer in the future within the next 2 to 3 years. At the moment, the use of desktop computers is very common because they are the traditional way of reading digital structured documents. In the future the percentage may highly decrease in favor of the emerging category of mobile and wearable devices like smart phones, smart tablets or smart watches.

The use of scanners in the future depends to 26 per cent on the use of desktop computers in the future, as shown by a moderate bivariate correlation with $v = 0.51$ and a very high level of statistical significance with $p = 4.31e-9$. Figure 4.54 shows the use of scanners in the future within the next 2 to 3 years in correlation to the use of desktop computers in the future within the next 2 to 3 years as well as the percentages of how often scanners will be used:

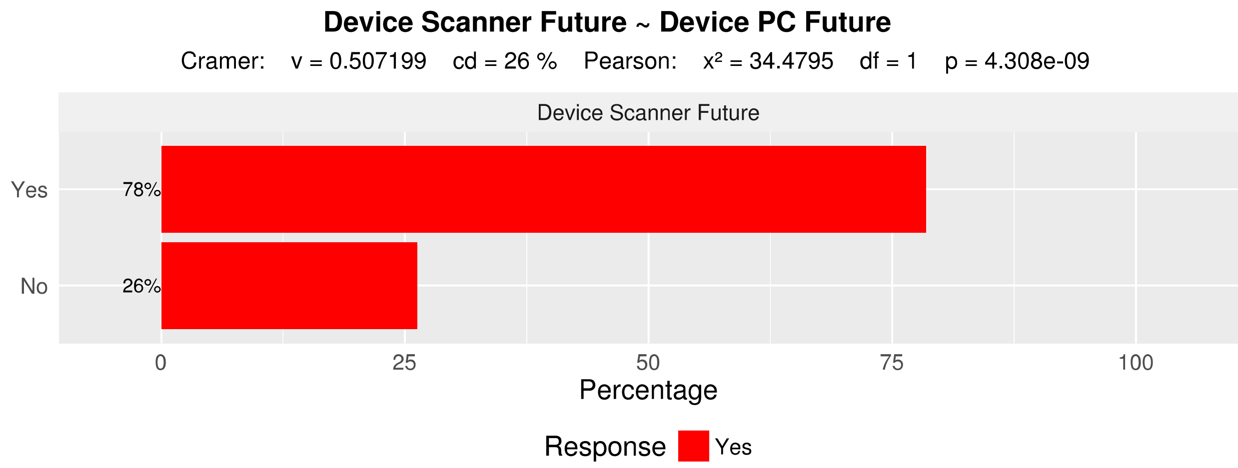


Figure 4.54: Scanners Future in correlation to Desktop Computers Future

78 per cent of the persons planning to purchase a new or additional desktop computer in the future are also intending to purchase a new or additional scanner in the future. On the other hand 74 per cent of the persons who do not plan to purchase a new or additional desktop computer in the future also have no plans to purchase an additional or new scanner in the future.

64 per cent of blind and visually impaired persons have a laptop computer and 50 per cent of them are intending to purchase a new or additional laptop computer in the future within the next 2 to 3 years.

58 per cent of the persons concerned have a mobile phone with camera and 43 per cent of them have plans to purchase a new or additional mobile phone with camera in the future within the next 2 to 3 years.

The use of mobile phones with camera in the future depends to 22 per cent on the use of mobile phones with camera today, as shown by a moderate bivariate correlation with v = 0.47 and a very high level of statistical significance with p = 5.58e-8. Figure 4.55 shows the use of mobile phones with camera in the future within the next 2 to 3 years in correlation to the use mobile phones with camera today as well as the percentages of how often mobile phones with camera will be used:

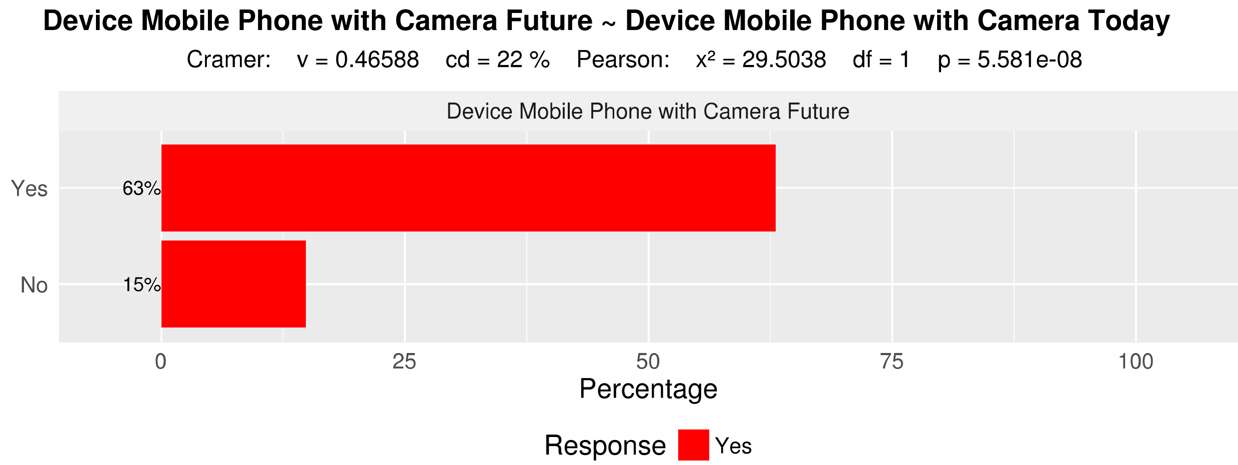


Figure 4.55: Mobile Phones with Camera Future to Mobile Phones with Camera Today

63 per cent of the persons using a mobile phone with camera today are also intending to purchase a new or additional mobile phone with camera in the future. On the other hand

85 per cent of the persons who do not use a mobile phone with camera today also have no plans to purchase an additional or new mobile phone with camera in the future.

Another moderate bivariate correlation with v = 0.48 and a very high statistical significance level of p = 3.04e-8 shows that the use mobile phones with camera in the future depends to 23 per cent on the use of mobile phones in the future in general. Figure 4.56 show the use of mobile phones with camera in the future within the next 2 to 3 years in correlation to the use of mobile phones in the future within the next 2 to 3 years in general as well as the percentages of how often mobile phones with camera will be used:

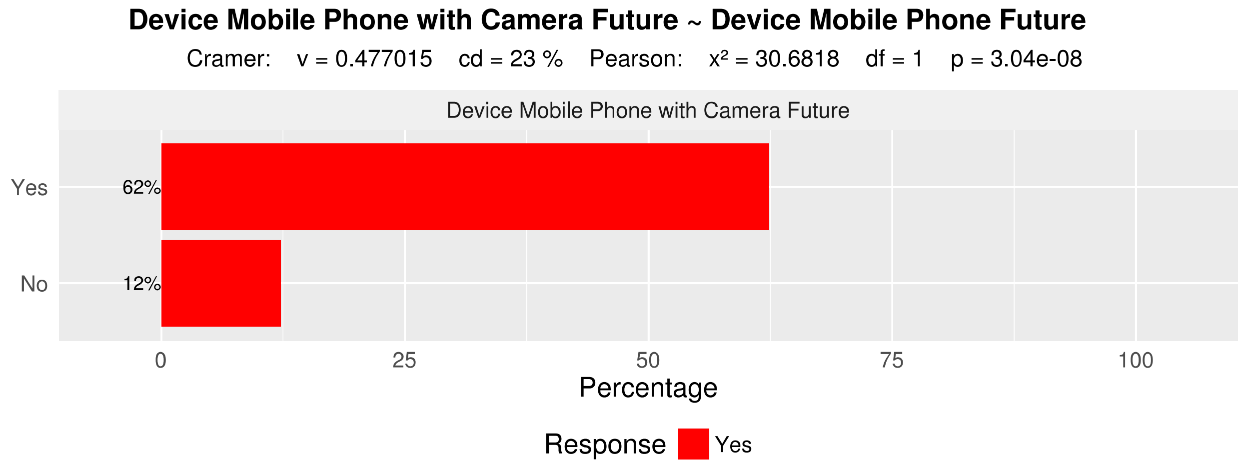


Figure 4.56: Mobile Phones with Camera Future to Use of Mobile Phones Future

62 per cent of the persons planning to purchase a new or additional mobile phone in the future are intending to purchase a mobile phone with camera. On the other hand 88 per cent of the persons who do not plan to purchase a new or additional mobile phone in the future also have no plans to purchase a mobile phone with camera in the future.

57 per cent of the respondents are using a Daisy player and 41 per cent of them are planned to purchase a new or additional Daisy player in the future within the next 2 to 3 years. The Daisy format is very popular with blind and visually impaired persons because it offers highly structured content on a combination of different media.

The use of Daisy players today is to 7 per cent dependent on the use of the acoustic physical medium, as shown by a weak bivariate correlation with v = 0.27 and a high level of statistical significance with p = 0.012. Figure 4.57 shows the use of Daisy players today in correlation to the use of the acoustic physical medium as well as the percentages of how often Daisy players are used:

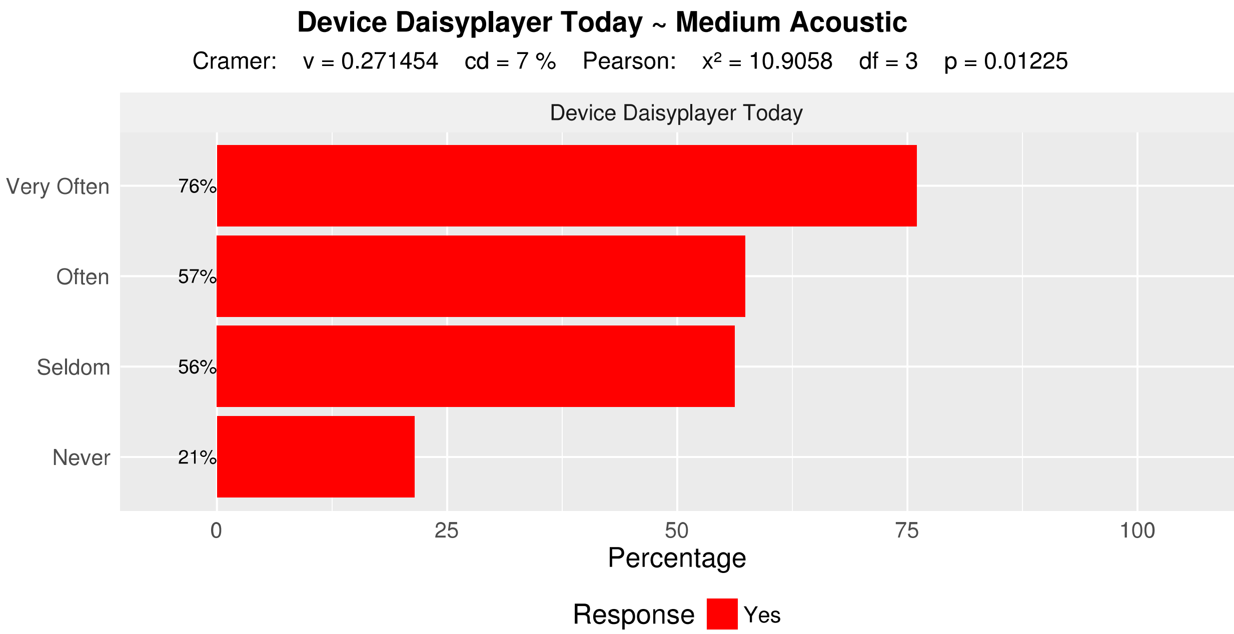


Figure 4.57: Daisy Players Today in correlation to the Acoustic Physical Medium

76 per cent of the persons who receive structured documents on the acoustic medium very often as well as 57 per cent of the persons receiving acoustic structured documents often, 56 per cent of the persons who receive structured documents on the acoustic medium seldom and only 21 per cent of the persons who never receive acoustic structured documents are using a Daisy player today.

Another moderate bivariate correlation with v = 0.48 and a very high statistical significance level of p = 1.95 shows that the use of Daisy players in the future depends to 23 per cent on the use of scanners in the future. Figure 4.58 shows the use of Daisy players in the future within the next 2 to 3 years in correlation to the use of scanners in the future within the next 2 to 3 years as well as the percentages of how often Daisy Players will be used:

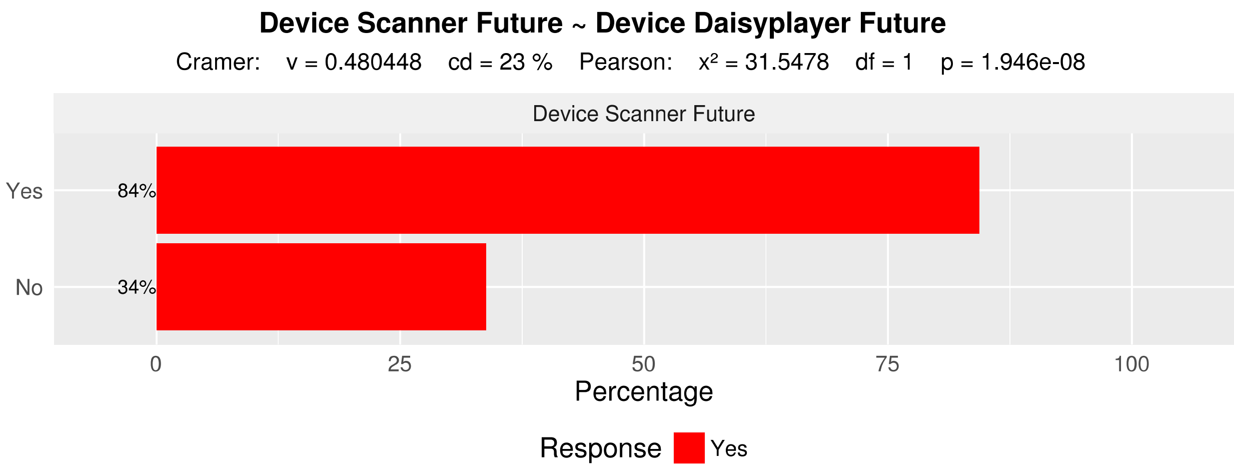


Figure 4.58: Daisy Players Future in correlation to Scanners Future

84 per cent of the persons planning to purchase a new or additional scanner in the future are also intending to purchase a new or additional Daisy player in the future. On the other hand 68 per cent of the persons who do not plan to purchase a new or additional scanner in the future also have no plans to purchase an additional or new Daisy player in the future.

A mobile phone with mobile Internet access is used by 37 per cent of the blind and visually impaired persons and 38 per cent of them are planning to purchase a new or additional mobile phone with mobile Internet access in the future within the next 2 to 3 years.

The use of mobile Internet access today depends to 23 per cent on the use mobile phones with camera today, as shown by a moderate bivariate correlation with v = 0.48 and a very high level of statistical significance with p = 2.99e-9. Figure 4.59 shows the use of mobile Internet access today in correlation to the use of mobile phones with camera today as well as the percentages of how often mobile Internet access is used:

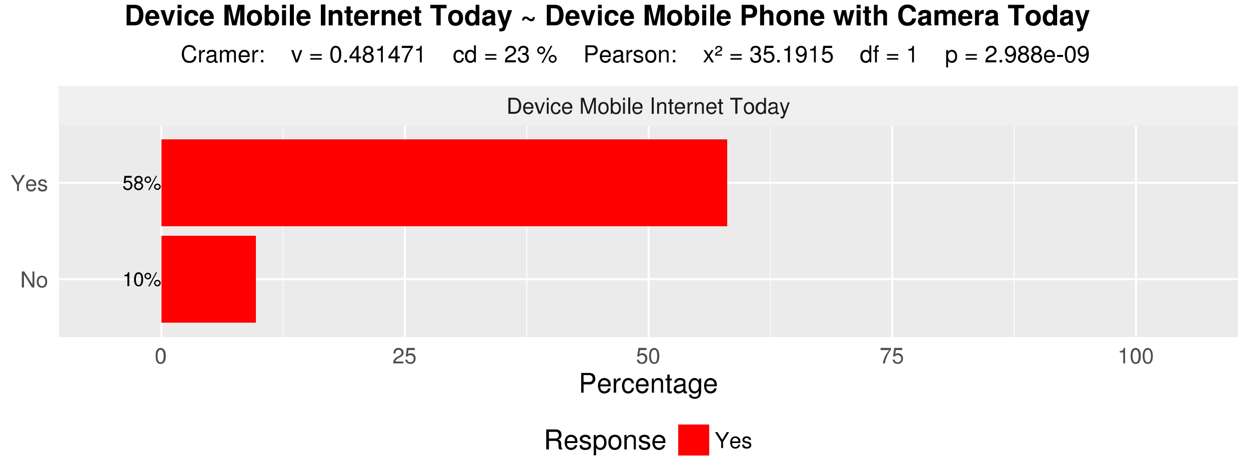


Figure 4.59: Mobile Internet Today in correlation to Mobile Phones with Camera Today

58 per cent of the persons who use a mobile phone with camera today are also using mobile Internet access today. On the other hand 90 per cent of the persons not using a mobile phone with camera today are also not using mobile Internet access today.

Another strong bivariate correlation with v = 0.62 and a very high statistical significance level of p = 9.75e-13 shows that the use of mobile Internet access in the future also depends to 39 per cent on the use of mobile phones with camera in the future. Figure 4.60 shows the use of mobile Internet access in the future within the next 2 to 3 years in correlation to the use of mobile phones with camera in the future within the next 2 to 3 years as well as the percentages of how often mobile Internet access will be used:

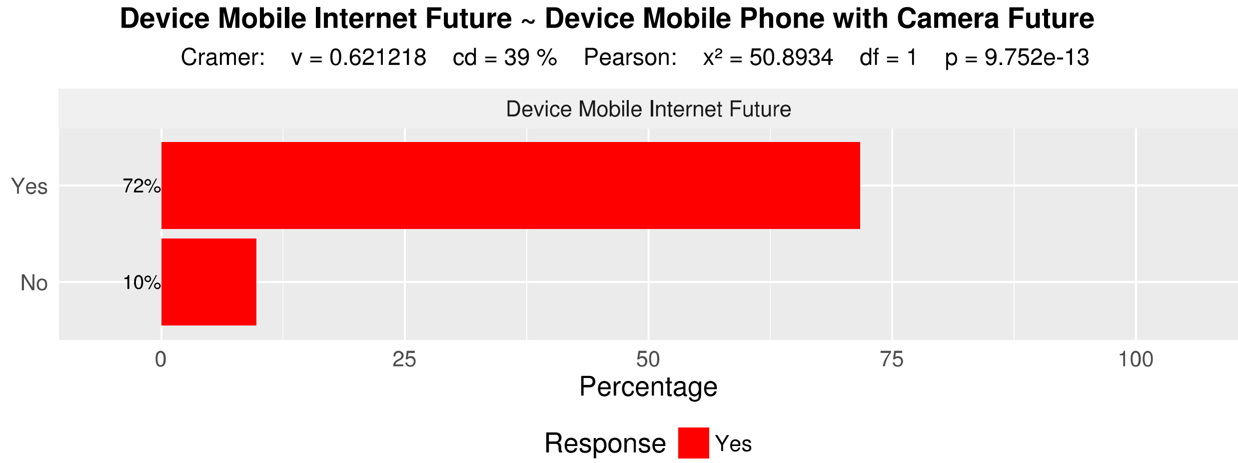


Figure 4.60: Mobile Internet Future to Mobile Phones with Camera Future

72 per cent of the persons planning to purchase a new or additional mobile phone with camera in the future are also intending to purchase mobile Internet access in the future.

On the other hand 90 per cent of the persons who do not plan to purchase a new or additional mobile phone with camera in the future also have no plans to purchase mobile Internet access in the future.

An E-Book reader is at the moment used by only 7 per cent of the persons concerned and 27 per cent of them have plans to purchase a new or additional E-Book reader in the future within the next 2 to 3 years. E-Book readers are used very seldom by blind and visually impaired persons because they offer limited possibilities for navigation since they are mainly purposed for the reading of novel books and not for the reading of highly structured documents. In addition the accessibility features of many E-Book reader products are limited at the moment.

The use of E-Book readers today depends to 22 per cent on the use of touch screen for desktop computers, as shown by a moderate bivariate correlation with v = 0.47 and a very high level of statistical significance with p = 6.19e-10. Figure 4.61 shows the use of E-Book readers today in correlation to the use of the touch screen input modality for desktop computers as well as the percentages of how often E-Book readers are used:

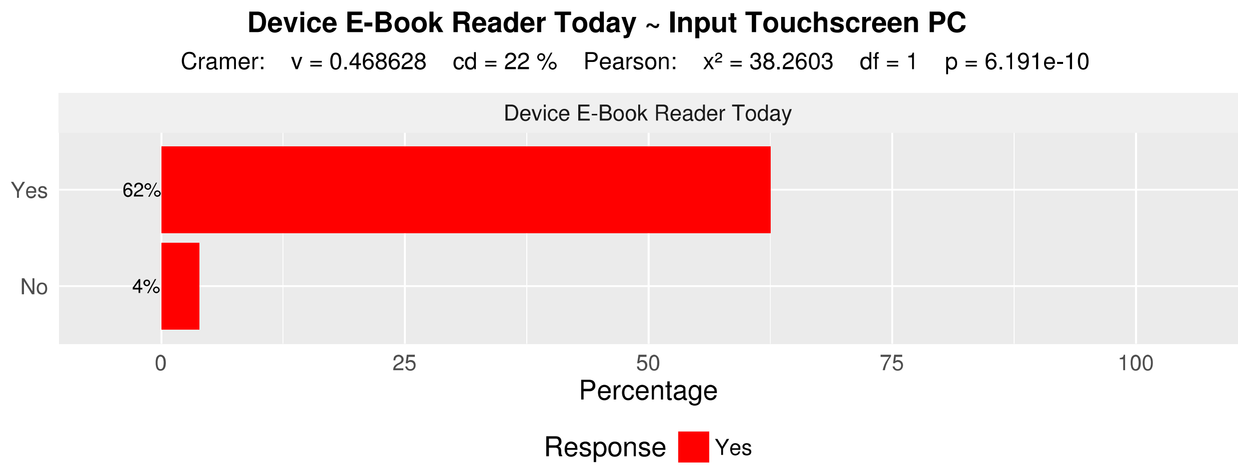


Figure 4.61: E-Book Readers Today to Touch Screen Input Modality for PC

62 per cent of persons using touch screen for desktop computers are also using an E-Book reader today. On the other hand 96 per cent of the persons who do not use touch screen for desktop computers are also not using an E-Book readers today.

Novel concepts should be developed for and implemented on the emerging category of mobile and wearable devices like smart phones, smart tablets or smart watches because this category of hardware devices are the devices most used by the persons concerned since blind and visually impaired persons would wish to have their assistive technology for the reading of structured document mobile more than bound to a fixed location.

4.4 Conclusions

In order to develop a novel concept for a multi-modal user interface for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices, a survey has been conducted among 205 blind and visually impaired persons to find out how they handle structured documents at the moment and what is of importance as to the reading of and the navigation within structured documents from the blind and visually impaired person's point of view.

There is serious need for action required by conducting further research on a novel concept for an assistive technology in support of blind and visually impaired persons for to the reading of structured documents because most of the persons concerned, namely

77 per cent of them, are not very satisfied with their current general situation as to the reading of structured documents according to the results 4.3.1 General Satisfaction.

Further research should be investigated in the digital physical medium because in most cases, blind and visually impaired persons would wish to receive their structured documents in digital form because of the emerging digital and web based platforms and services on which structured documents are provided to them and since digital structured documents are directly accessible to a variety of different assistive technologies and can easily be adapted according to the special needs according to the results 4.3.2 Physical Media.

A novel concept for an assistive technology in support of blind and visually impaired persons for the reading of structured documents should be format independent in order to be able to interworking with all the different digital content formats available out there because structured documents are provided to blind and visually impaired persons in a mix of different digital content formats according to the results 4.3.3 Digital Content Formats.

More than one human sense should be served by a novel concept for an assistive technology in support of blind and visually impaired persons for the reading of structured documents because in most cases, a combination of multiple assistive technologies employing different input and output modalities are used simultaneously at the same time in order to serve the special needs on multiple human senses like the visual, the aural or the tactile sense according to the results 4.3.4 Assistive Technologies.

Novel input modalities like gestures on multi-touch screen or multi-touch pad, motion or speech, provided by the emerging category of mobile and wearable devices out of the box, should be employed in addition to the traditional input modalities like keyboard and mouse in a novel concept for an assistive technology in support of blind and visually impaired persons for the reading of structured documents because the persons concerned would wish to have a more easy, intuitive and standardised way for the reading of and the navigation within structured documents according to the results 4.3.5 Input Modalities.

The logical document structure plays a very important role as to the reading of and the navigation within structured documents. A novel concept should be developed for non-visual presentation of the logical structure elements employing novel output modalities like auditory icons or vibration feedback, provided by the emerging category of mobile and wearable devices out of the box, in addition to the traditional output modalities like speech synthesis or Braille displays used by state of the art screen readers because all of the participants would wish to have a better overview over the logical document structure of structured documents according to the results 4.3.6 Structure Elements.

At the moment most structured documents available on the Internet are not or not fully accessible for blind and visually impaired persons because all of the persons concerned reported that they receive documents which they cannot access. This is not a problem of missing scientific concepts more than that these concepts did not move into the praxis. To improve the current situation authors of structured documents available on the Internet like web developer, web designer and web redactors should be highly encouraged to follow certain accessibility standards and tools like plug-ins for authoring tools should be developed for checking the accessibility of a document and helping the author to fixing certain accessibility issues according to the results 4.3.7 Accessibility Problems and 4.3.8 Accessibility Solutions.

Novel concepts should be developed for and implemented on the emerging category of mobile and wearable devices like smart phones, smart tablets or smart watches because this category of hardware devices are the devices most used by the persons concerned since blind and visually impaired persons would wish to have their assistive technology for the reading of structured document mobile more than bound to a fixed location according to the results 4.3.9 Hardware Devices.

5 DOKY: A Multi-Modal User Interface for Non-Visual Presentation, Navigation and Manipulation of Structured Documents on Mobile and Wearable Devices

5.1 Introduction

The following multi-modal user interface for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices like smart phones, smart watches or smart tablets has been developed as a result of the inductive research among 205 blind and visually impaired participants in the previous chapters. It enables the user to get a fast overview over the document structure and to efficiently skim and scan over the document content by identifying the type, level, position, length, relationship and content text of each element as well as to focus, select, activate, move, remove and insert structure elements or text. These interactions are presented in a non-visual way using Earcons, Tactons and synthetic speech utterances, serving the auditory and tactile human sense. Navigation and manipulation is provided by using the multitouch, motion (linear acceleration and rotation) or speech recognition input modality. It is a complete solution for reading, creating and editing structured documents in a non-visual way. There is no special hardware required. The name DOKY is derived from a short form of the terms document and accessibility.

5.2 User Interface Design

The user interface design defines how information is presented to the used and which controls the user interface provides for the user to give input and to performing actions. Figure 5.1 gives an overview over the user interface design employed in this research:

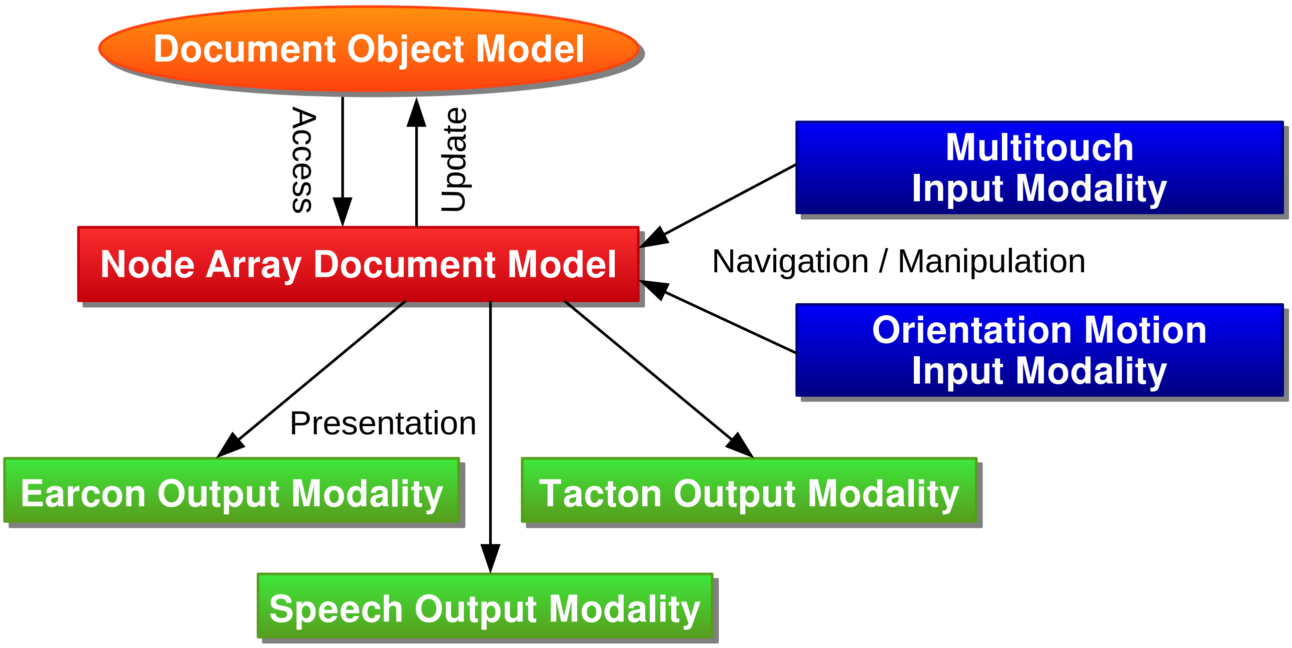


Figure 5.1: User Interface Design Overview

The structured document is provided to the user interface through the W3C Document Object Model (DOM) (Van Kesteren, Aryeh, Russell & Berjon, 2015). This is a platform and language neutral application programming interface which allows the user interface to dynamically access and update the document structure and content. In the DOM, documents are represented as a hierarchy of element and text node objects which is very much like a tree. The W3C Document Object Model (DOM) has been introduced in detail in Chapter 2.4.1.

In the next step, the document is reorganised and transformed into the Node Array Document Model, which is an alternative modality-neutral document model for presentation, navigation and manipulation of a structured documents. In contrast to the tree-like document representation employed in the Document Object Model described above, the element and text objects of the document are organised in a two-dimensional array.

The multi-touch input modality allows the user to perform the actions for navigation and manipulation provided by the node array model by performing multi-touch gestures on the multi-touch screen or multi-touch pad of the mobile and wearable device and the orientation motion input modality by performing motion gestures by moving the mobile or wearable device itself.

Each interaction of the node array model is presented to the user in a non-visual way by the earcon, tacton and speech output modality, serving the auditory and the tactile human sense.

5.2.1 Node Array Document Model

The Node Array Document Model is an alternative document model for representing structured documents which provides functions for navigation and manipulation. According to Van Kesteren, Aryeh, Russell & Berjon (2015), a document model is a collection of descriptions of data structures and their contained fields, together with the operations or functions that manipulate them. Figure 5.2 shows the controls and functions of the node array document model on an example document:

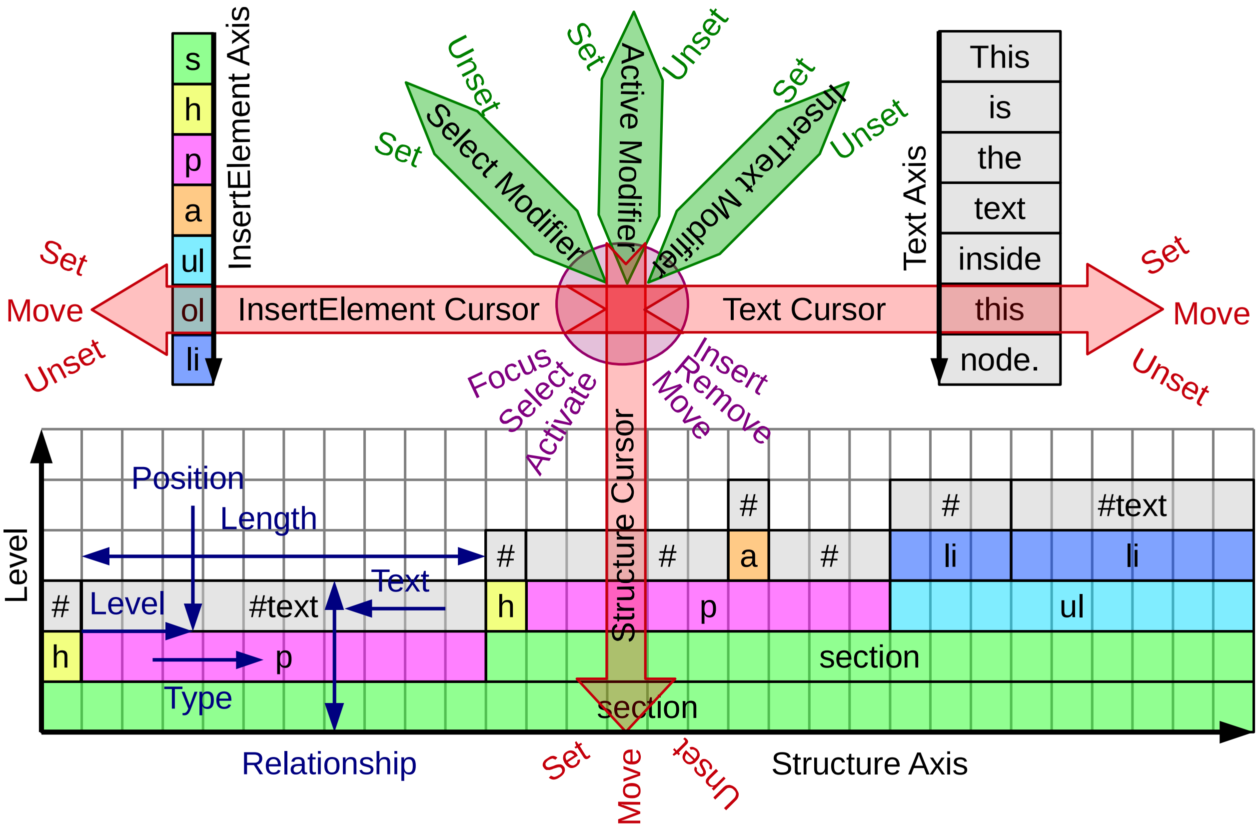


Figure 5.2: Node Array Document Model Overview on an Example Document

The element and text objects of the structured document are arranged and organised in a two dimensional array. On each hierarchy level, the element and text objects of the document structure are laid-out along the structure axis in document reading order where the position on the axis represents the position of the object within the document and the length on the axis represents the length of the text contained in it. The advantages of this representation is that is enables the user to get a fast overview over the document structure and to efficiently skim and scan over the document content by identifying the type, level, position, length, relationship and text of each element. For each text node there is an additional text axis on which the contained text is laid out by word. The element types which can be inserted at a specific position are laid out along an addition insert element axis.

There is a structure, text and insertElement cursor. Each cursor can be set to a specific position, moved forward and backward at an arbitrary speed and unset. There is a select, active and insertText modifier. Each modifier can be set or unset.

By performing the actions provided by the cursors and modifiers different operations can be performed on the element and text objects. The operations are focus, select, activate, move, remove and insert. These interactions are described in more detail in the user interaction design in Chapter 5.3.

5.2.2 Earcon Output Modality

The earcon output modality presents the interactions of the node array document model using earcons (Blattner, Sumikawa & Greenberg, 1989), which are non-verbal audio messages that are used in the computer user interface to provide information to the user about some computer object, operation or interaction. They use abstract, synthetic tones in structured combinations to create auditory messages. Earcons had been introduced in detail in Chapter 2.2.2.

For creating the set of earcons in this user interface the guidelines which have been drawn up by Brewster, Wright & Edwards (1992) and (1993) to help designers when creating earcons had been used along with the more general guidelines for the syntactic design and integration of audio cues in computer user interfaces given by Sumikawa (1985) and Sumikawa, Blattner, Joy & Greenberg (1986). They allow a designer with no knowledge of sound design to create a set of earcons that will effectively communicate complex information in sound. He or she can be sure that the earcons will be perceivable and recognisable by listeners because they incorporate knowledge of earcon design and psychoacoustics.

Each interaction is presented by a single-pitch inherited elementary earcon inheriting the timbre of the element type, pitch of the element level, spatial location of the element position, register of the operation and dynamics of the status. A single-pitch earcon is any audio message composed of only one note. Inherited, family or hierarchical earcons provide a powerful, hierarchical system. Each earcon is a node on a tree and inherits all the properties of the earcons above it. With the inheritance hierarchy there is only one new piece of information to remember at each level, thus making family earcons easier to remember. Initially, there will be much to learn but later extensions will be simple.

These single-pitch inherited elementary earcons are rendered simultaneously as parallel compound earcons for all interactions that are going on at the same time. Parallel compound earcons proposed by Brewster, Wright & Edwards (1993) are a method to reduce the length of time a compound audio message takes by playing the sequential component parts of a compound earcon in parallel so that they only take the time of one to play. Blattner, Papp & Glinert (1992) say that our awareness and comprehension of the auditory world around us for the most part is done in parallel. This suggests that parallel compound earcons could exploit a natural ability of the human auditory system. Parallel compound earcons also use some of the attributes of the musical theory of counterpoint defined by Scholes (1975). In counterpoint individual instruments play separate musical lines which come together to make a musical whole. With parallel earcons, each component earcon is separate but the whole combined sound gives the meaning. Another factor that may give parallel earcons an advantage is the recency effect (Badeley, 1990), the term used to describe the enhanced recall of the most recently presented items. Parallel earcons have all parts played at the same time so there is less time to forget any one earcon.

Element Type

The element type of the element node associated with the interaction and on which the operation of the interaction is performed is presented by the timbre of the single-pitch inherited elementary earcon. Blattner, Sumikawa & Greenberg (1989) suggest the use of simple timbres such as sine, square, sawtooth and triangular wave but psychoacoustics (the study of the perception of sound) suggests that complex musical instrument timbres may be more effective and easily recognised (Moore, 1989) than simple tones. This is due to their greater spectral and temporal complexity making them more discriminable than simple tones. Brewster, Wright & Edwards (1993) also proposes to use complex musical instrument timbres. Where possible timbres with multiple harmonics should be used as this helps perception and can avoid masking. Timbres should be used that are subjectively easy to tell apart. For example, on a musical instrument synthesiser brass and organ rather than brass1 and brass2. However, instruments that sound different in real life may not when played on a synthesiser, so care should be taken when choosing timbres.

According to these guidelines, the following 10 standard element types of structured text-documents had been used in this research and mapped to the following complex musical instrument timbres as shown in Table 5.1. However, this table is not exhaustive. It can be extended by adding more element types of structured text-documents or any other type of structured documents. Alternatives were not considered.

Element Type	Complex Musical Instrument Timbre
Section	Brass
Heading	Trumpet
Paragraph	Guitar
Graphic	Glockenspiel
Unordered List	French Horn
Ordered List	English Horn
List Item	Oboe
Table	Organ
Table Row	Violin
Table Cell	Clarinet

Table 5.1: Element Types to Complex Musical Instrument Earcon Timbres Mapping

As an alternative approach to complex musical instrument timbres spearcons or speech-based earcons (Walker, Nance & Lindsay, 2006) may be employed. These are brief audio cues that can play the same roles as earcons and auditory icons, but in a more effective manner. Spearcons had been discussed in detail in Chapter 2.2.3.

The advantages of spearcons are that they can be created automatically by converting the text of the name of an element type, for example “Heading”, to speech via text-to-speech (TTS), and then speeding up the resulting audio clip (without changing pitch) to the point that it is no longer comprehensible as speech. Spearcons are unique to the specific element type. These unique sounds are analogous to fingerprints because of the acoustic relation between the spearcons and the original speech phrases. At the same time, the similarities in element types cause the spearcons to form families of sounds. For example, the spearcons for “Table”, “Table Row” and “Table Cell” are all unique, including being of different lengths. However, they are acoustically similar at the beginning of the sounds, which allows them to be grouped together even though they are not comprehensible as any particular words.

Since the mapping between spearcons and their element types is non-arbitrary, there is less learning required than would be the case for a purely arbitrary mapping as with the complex musical instrument timbres described above. Spearcons provide more direct mappings between sound and element type than musical instrument timbres, and cover more content domains, more flexibly. Spearcons can be created algorithmically, so they can be created dynamically, and can represent any possible element type. Also, spearcons are easy to learn whether they are comprehensible as a particular word or not, because they derive from the original speech (Palladino & Walker, 2007).

This has not been tested in this research. A further test would be required to investigate if spearcons are an effective method for presenting the element type to the user.

Level

The level of the element node associated with the interaction and on which the operation of the interaction is performed is presented by the pitch of the single-pitch inherited elementary earcon. Sumikawa (1985) suggest that notes should be kept within a range of eight octaves of twelve notes. Sumikawa, Blattner, Joy & Greenberg (1986) also suggest that random combinations of pitches should not be used, but that they should be taken from one octave and in the same scale for easier manipulation. This fits in well with the work of Dewar, Cuddy & Mewhort (1977) who suggested that listeners can better detect differences between groups of sounds if all the notes in one sound are in a different scale to the other. Keeping all the notes in one octave also minimises pitch perception problems where a listener can mistake the octave to which a note belongs (Deutsch, 1986). Semitone gaps should be avoided as they can create incorrect melodic implications. Earcons should be musically neutral. Brewster, Wright & Edwards (1993) suggests that pitch should not be used on its own unless there are large differences between those used (see register below). Complex intra-earcon pitch structures are effective in differentiating earcons if used along with rhythm. Some suggested ranges for pitch are maximum 5 kHz (four octaves above middle C) and minimum 125 Hz - 150 Hz (the octave of middle C).

According to these guidelines, the following levels had been used in this research and mapped to the following pitches as shown in Table 5.2. This gives a maximum of 12 possible document levels. Alternatives were not considered.

Level	Pitch
1	C
2	C#
3	D
4	D#
5	E
6	F
7	F#
8	G
9	G#
10	A
11	A#
12	B

Table 5.2: Node Levels to Earcon Pitches Mapping

Position

The position of the element node associated with the interaction and on which the operation of the interaction is performed is presented by the spatial location of the single-pitch inherited elementary earcon.

Another important factor in the perception of sound is spatial location. This is the ability to identify the position of a sound source in space. If a sound source is located to one side of the head, then the sound reaching the further ear will be reduced in intensity due to interaural intensity difference (IID) and delayed in time due to interaural time difference (ITD). This is known as the Duplex Theory and was first developed by Lord Rayleigh in 1907 as described by Moore (1989) and Levitt & Voroba (1974). IID and ITD can be used in auditory interfaces to provide directional or positional information.

Moore (1989) suggests that there can be an interaural intensity difference (IID) of up to as much as 20 dB across the ears. It can be seen that IID has little effect at low frequencies but is much more important for localisation at higher frequencies above 3 kHz. Interaural time difference (ITD), on the other hand, is only useful for localisation at lower frequencies because as frequencies increase to above 1500 Hz, the wavelength is shorter than the distance between the ears. When the sound source is at 90 degrees to the head (opposite one ear), there is a greater than 0.6 milliseconds delay between the ears. Delays of up to 2 milliseconds can be heard but longer than this the sound tends to be heard as two separate tones. Scharf & Houtsma (1986) provides more details on IID and ITD.

Humans can detect small changes in the position of a sound source. The minimum auditory angle (MAA) is the smallest separation between two sources that can be reliably detected. Strybel, Manligas & Perrott (1992) suggest that in the median plane sound sources only 1 degree apart can be detected. The median plane cuts through the head vertically, along the line of the nose. At 90 degrees azimuth (directly opposite one ear) sources must be 40 degrees apart. This has important implications for auditory displays. It indicates that high-resolution sounds can be used when presented in front of the user.

Much of the recorded sound we hear is in stereo. As Burgess (1992) describes, along with the sound, a stereo recording captures differences in intensity. The perceived position is along a line between the speakers. This simple, inexpensive technique can give useful spatial cues at the auditory interface. Pitt & Edwards (1991) have shown that using IID a user can find targets on an auditory display accurately and with reasonable speed. A listener wearing headphones uses lateralisation to locate the position of a sound which will be perceived as being on a plane between the ears and within the head (Burgess, 1992). Work has been done by Sakamoto, Gotoh & Kimaura (1976), Wenzel, Foster, Wightman & Kistler (1989), Gerhing & Morgan (1990) and Wenzel, Wightman & Kistler (1991) to make sounds perceived through headphones appear fully in three dimensions, or outside the head.

According to these guidelines, in this research, the position of an element or text on the structure axis of the node array document model is mapped to the spatial location in stereo space along a line between the speakers from left to right. Alternatives were not considered.

Operation

The operation of the interaction which is performed on the element node associated with the interaction is presented by the register of the single-pitch inherited elementary earcon. The register is the position of the motive in the musical scale. A high register means a high pitched note and a low register a low pitched note. The same motive in a different register can convey a different meaning. Sumikawa (1985) suggest the only three registers (low, medium and high) should be used. Brewster, Wright & Edwards (1993) proposes that if listeners must be able to make absolute rather than relative judgements of earcons then pitch and register should not be used. A combination of pitch and another parameter would give better performance. If register alone must be used then there should be large differences between earcons but even then it might not be the most effective method. Two or three octaves difference should be used. This is not a problem in this research because only relative judgements between the different operations are to be made.

According to these guidelines, the following 6 operations of an interaction had been used in this research and mapped to the following different registers as shown in Table 5.3. Alternatives were not considered.

Operation	Register
FOCUS	6
SELECT	5
ACTIVE	4
MOVE	3
REMOVE	2
INSERT	7

Table 5.3: Interaction Operations to Earcon Registers Mapping

Status

The status of an interaction is presented by the dynamics of the single-pitch inherited elementary earcon. This is the change in volume of the note. It can be made to increase as the note plays (crescendo) or decrease (decrescendo). Sumikawa (1985) suggest a total of five dynamics (soft, medium, loud, soft to loud and loud to soft). Brewster, Wright

& Edwards (1993) says that care must be taken in the use of intensity. The overall sound level will be under the control of the user of the system. Earcons should all be kept within a close range so that if the user changes the volume of the system no sound will be lost. Some suggested ranges from Patterson (1982) are maximum 20 dB above threshold and minimum 10 dB above threshold.

One important ability of the human auditory system is that of auditory habituation (Buxton, Gaver & Bly, 1991) where continuous sounds with a restricted loudness range can fade into the background of consciousness after a short period of time. If the sound was to change or stop then it would come to the foreground of attention because of the sensitivity of the auditory system to changes in stimulus (Buxton, 1989).

According to these guidelines, the following 3 status of an interaction had been used in this research and mapped to the following different dynamics as shown in Table 5.4. Alternatives were not considered.

Status	Dynamics
STARTED	• Decrescendo from 20 dB to 10 dB within 0.0825 seconds • Afterwards continues at 10 dB
COMPLETED	• Crescendo from 10 dB to 20 dB within 0.0825 seconds • Afterwards stopped at 0 dB
ABORTED	• Stopped immediately at 0 dB

Table 5.4: Interaction Status to Earcon Dynamics Mapping

When a new interaction is started, the corresponding single-pitch inherited elementary earcon starts to play at a volume of 20 dB above threshold. It is than made to fade into the background of consciousness by lowering its volume from 20 dB to 10 dB above threshold within 0.0825 seconds. Brewster, Wright & Edwards (1993) says that small note lengths might not be noticed so notes less than sixteenth notes or semi-quavers should not be used. This depends on the tempo. If 180 beats per minute are used then sixteenth notes last 0.0825 seconds. Afterwards it is played continuous at a volume of 10 dB above threshold as long as the interaction is going on to achieve auditory habituation.

When an interaction is completed, the corresponding single-pitch inherited elementary earcon is made to come to the foreground of attention again by increasing its volume from 10 dB to 20 dB above threshold within 0.0825 seconds. Afterwards it stops playing by decreasing its volume from 20 dB to 0 dB above threshold.

On the other hand, when an interaction is aborted, the corresponding single-pitch inherited elementary earcon stops playing immediately by decreasing its volume from 10 dB to 0 dB above threshold.

5.2.3 Tacton Output Modality

The tacton output modality presents the interactions of the node array document model using tactons or tactile icons (Brewster & Brown, 2004). These are structured, abstract vibrotactile messages that can be used for presenting multidimensional information non-visually. Tactons had been introduced in detail in Chapter 2.2.5.

Tactons are created by encoding information using the parameters of cutaneous perception. Cutaneous perception refers to the mechanoreceptors contained within the skin, and includes the sensations of vibration, temperature, pain and indentation. Tactile devices are used to present feedback to the cutaneous sense. As tactons are abstract the mapping between the tacton and what it represents must be learned, but work on earcons has shown that learning can take place quickly (Brewster, 1998).

The encoding is similar to that of earcons in sound (Blattner, Sumikawa & Greenberg, 1989) where each of the musical parameters, for example timbre, pitch or register is varied to encode information. Similar parameters can be used for tactons although their relative importance is different. The properties that can be manipulated for tactons are similar to those used in the creation of earcons. The parameters for manipulation also vary depending on the type of transducer used. Not all transducers allow all types of parameters to be manipulated.

According to Brown (2007), mobile and wearable devices commonly have a very simple single point-contact tactile stimulator built-in that can alert the user to a call. These are small DC motors featuring an eccentric weight in the shaft. When they are switched on the shaft spins, and the spinning of the eccentric weight causes the sensation of vibration. These vibration motors typically vibrate at frequencies around 130 Hz. With a phone motor the whole casing in which the motor is housed vibrates. Therefore, when this motor is mounted in a mobile or wearable device, the whole body of the device vibrates. In addition, less control over the stimuli is available. Only the on-off times and the velocity of the rotation can be controlled. This enables them to produce pulses of different durations and intensities. Therefore, when using phone vibration motors, tactons must be designed within these constraints. Building on from duration, groups of pulses of different durations can be composed into rhythmic units and the tempo at which these rhythms are played can also be varied.

Each interaction is presented by a single-motive inherited elementary tacton inheriting the rhythm of the element type, the tempo of the element level and the intensity of the operation. As suggested by Blattner, Sumikawa & Greenberg (1989), short motifs could be used to represent simple objects or operations and these can then be combined in different ways to represent more complex messages and concepts. As with inherited, family or hierarchical earcons, tactons can also be combined in a hierarchical way. Each tacton is a node in a tree and inherits properties from the levels above it. With the inheritance hierarchy there is only one new piece of information to remember at each level, thus making family tactons easier to remember. Initially, there will be much to learn but later extensions will be simple.

These single-motive inherited elementary tactons are rendered sequentially as serial compound tactons for all interactions that are going on at the same time. Serial compound tactons are the simplest type of complex tactile messages and they are formed by placing two or more elementary tactons in succession. This provides a simple and effective method for building up complex messages in vibration.

Element Type

The element type of the element node associated with the interaction and on which the operation of the interaction is performed is presented by the rhythm of the single-motive inherited elementary tacton. Rhythm is a very effective parameter in both earcon design and tactons. Using different rhythms can create a large number of distinguishable motifs. Deutsch (1980) says that rhythm is one of the most important characteristics of a sound and Blattner, Sumikawa & Greenberg (1989) describe this as the most prominent characteristic of a motive. Sumikawa (1985) suggested that only seven time divisions should be used when creating rhythms. She also said that motives should be no longer than three or four notes or they will become too long for the user to easily remember and may also take too much time to play when in combination. Brewster, Wright & Edwards

(1993) suggested the rhythms should be made as different as possible. Putting different numbers of notes in each rhythm is very effective. Patterson (1982) says that sounds are likely to be confused if the rhythms are similar even if there are large spectral differences. Small note lengths might not be noticed so notes less than sixteenth notes or semi-quavers should not be used. This depends on the tempo. If 180 beats per minute is used then sixteenth notes last 0.0825 seconds.

According to these guidelines, the following 10 standard element types of structured text-documents had been used in this research and mapped to the following rhythms as shown in Table 5.5. However, this table is not exhaustive. It can be extended by adding more element types of structured text-documents or any other type of structured documents. Alternatives were not considered.

Element Type	Rhythm
Section	Three eighth notes
Heading	One eighth and one full note
Paragraph	One quarter and one eighth note
Graphic	One full one eighth and one half note
Unordered List	One full and one half note
Ordered List	One full and one quarter note
List Item	One quarter one half and one eighth note
Table	One quarter one half and one full note
Table Row	One quarter one eighth and one quarter note
Table Cell	One quarter one eighth and one full note

Table 5.5: Element Types to Tacton Rhythms Mapping

Level

The level of the element node associated with the interaction and on which the operation of the interaction is performed is presented by the tempo of the single-motive inherited elementary tacton.

Brewster, Wright & Edwards (1995) said that changing the tempo, speeding up or slowing down the sounds, is another effective method for differentiating earcons. In addition to following Brewster's guidelines also advice given by Van Erp & Spapé (2003) who identified tempo (speed) as an important parameter in the identification of tactile melodies has been followed. The tempo is expressed in beats per minute (bpm).

According to these guidelines, the following levels had been used in this research and mapped to the following tempos as shown in Table 5.6. This gives a maximum of 12 document levels. Alternatives were not considered.

Level	Tempo
1	180 bpm
2	195 bpm
3	210 bpm
4	225 bpm
5	240 bpm
6	255 bpm
7	270 bpm
8	285 bpm
9	300 bpm
10	315 bpm
11	330 bpm
12	345 bpm

Table 5.6: Node Levels to Tacton Tempos Mapping

Operation

The operation of the interaction which is performed on the element node associated with the interaction is presented by the intensity of the single-motive inherited elementary tacton. Brown (2007) says that, when using a phone vibration motor, the intensity of the vibration is changed by adjusting the supply voltage to the phone motor. When the supply voltage of a phone motor is adjusted it changes both the amplitude and the frequency of the resulting vibration. This, therefore, provides redundant coding of the same information in both of these parameters of vibration and may help people to identify multiple levels. Sherrick & Cholewiak (1986) found that combining frequency and amplitude redundantly allowed a greater number of identifiable levels to be created.

A range of frequencies can be used to differentiate tactons. The range of 20 – 1000 Hz is perceivable but maximum sensitivity occurs around 250 Hz (Gunther, Davenport & O'Modhrain, 2002). The number of discrete values that can be differentiated is not well understood, but Gill (2003) suggests that a maximum of nine different levels can be used. As in audition, a change in amplitude leads to a change in the perception of frequency so this has an impact on the use of frequency as a cue. The number of levels of frequency that can be discriminated also depends on whether the cues are presented in a relative or absolute way. Making relative comparisons between stimuli is much easier than absolute identification, which will lead to much fewer discriminable values, as shown in the work on earcon design (Brewster, Wright & Edwards, 1992).

The amplitude of stimulation can also be used to encode values to present information to the user. Gunther, Davenport & O'Modhrain (2002) reports that the intensity range of the skin extends to 55 dB above the threshold of detection, beyond which vibrations may become unpleasant or painful (Vitense, Jacko & Emery, 2003). Craig & Sherrick (1982) indicate that perception deteriorates above 28 dB so this would seem to be a useful maximum. Gunther (2001) reports that various values, ranging from 0.4 dB to 3.2 dB, have been reported for the just noticeable difference (JND) value for intensity. Gill (2003) states that that no more than four different intensities should be used. Again the number of useful discriminable values will depend on absolute or relative presentation of stimuli.

Due to the interactions between frequency and amplitude several researchers have suggested that they be combined into a single parameter to simplify design.

According to these guidelines, the following 6 operations of an interaction had been used in this research and mapped to the following different intensities as shown in Table 5.7. Alternatives were not considered.

Operation	Intensity
FOCUS	100 %
SELECT	95 %
ACTIVE	90 %
MOVE	85 %
REMOVE	80 %
INSERT	75 %

Table 5.7: Interaction Operations to Tactons Intensities Mapping

5.2.4 Speech Output Modality

The speech output modality presents the interactions of the node array document model using synthetic speech. Synthetic speech had been introduced in detail in Chapter 2.2.4.

Presenting information in speech is slow because of its serial nature. To assimilate information the user must hear it from beginning to end and many words may have to be comprehended before a message can be understood. Speech suffers from many of the same problems as text in text-based computer systems, as this is also a serial medium.

Synthetic speech is mainly purposed for presenting text. However by manipulating attributes of the speech like the voice, pitch register and spatial location it is also possible to encode some non-text data attributes along with a text.

Each interaction is presented by a inherited elementary utterance, inheriting the voice of the element type, pitch of the node level, register of the operation and text of the content. As suggested by Blattner, Sumikawa & Greenberg (1989), short utterances could be used to represent simple objects or operations and these can then be combined in different ways to represent more complex messages and concepts. As with inherited, family or hierarchical earcons and tactons, speech can also be combined in a hierarchical way. Each utterance is a node in a tree and inherits properties from the levels above it. With the inheritance hierarchy there is only one new piece of information to remember at each level, thus making family tacton easier to remember. Initially, there will be much to learn but later extensions will be simple.

These inherited elementary utterances are rendered sequentially as serial compound utterances for all interactions that are going on at the user interface at the same time. Serial compound utterances are the simplest type of complex speech messages and they are formed by placing two or more elementary utterances in succession. This provides a simple and effective method for building up complex messages in speech.

Element Type

The element type of the element node associated with the interaction and on which the operation of the interaction is performed could be presented by the voice of the inherited elementary utterance.

Sorin, Lemarié, Aussenac-Gilles, Mojahid & Oriola (2014) used specialized audio and two voices to demarcate headings in their work on communicating text structure to blind persons with text-to-speech. Their research showed that text comprehension was slightly improved but failed to show statistically significant evidences that either the dual-voices method or the spatialised audio method has better performance than current text-to-speech text description saying “Heading level N” before the heading oralisation. However, their document structures were very simple, only consisting of two different element types, headings and non-heading document contents. If more complex document structures were used then a performance increase may have been found.

This has not been tested in this research. A further test would be required to investigate if different voices are an effective method for presenting the element type to the user.

Level

The level of the element or text node associated with the interaction and on which the operation of the interaction is performed is presented by the pitch of the inherited elementary utterance. In this research, the same pitches are used as with the earcon output modality as described in detail in Chapter 5.2.2 and in Table 5.2.

Position

The position of the element or text node associated with the interaction and on which the operation of the interaction is performed is presented by the spatial location of the inherited elementary utterance. In this research, the same spatial locations are used as with the earcon output modality as described in detail in Chapter 5.2.2.

Operation

The operation of the interaction which is performed on the element or text node associated with the interaction is presented by the register of the inherited elementary utterance. In this research, the same registers are used as with the earcon output modality as described in detail in Chapter 5.2.2 and in Table 5.3.

Text

The text of the element or text node associated with the interaction and on which the operation of the interaction is performed is presented by the text of the inherited elementary utterance.

If the node is a text node, than the contained text is used. If the node is an element node and contains a heading element, the text of this heading element is used instead.

This text is presented at a speaking rate of 150 words per minute, which is far slower than the maximum listening speed for blind persons using Text-To-Speech (Asakawa & Takagi, 2003). Slowiaczek & Nusbaum (1985) found that a speaking rate of about 150 words per minute is optimal for perception of synthetic speech. This figure is around the normal speaking rate and is very slow to listen to. When they increased the rate to 250 words per minute (the normal sight-reading rate) recognition decreased significantly. One of the main causes of this, they suggest, is the poor quality of the synthetic speech. Much of the prosodic information (intonation, pausing etc.) in normal speech is not given in synthetic speech. At low speeds listeners can cope without it but at higher speeds it becomes much more important. With these problems, users of synthetic speech will be constrained to operate at rates that are far below normal reading rates. Highly-skilled users, such as blind and visually impaired persons, can reach higher recognition rates but this requires much practice.

5.2.5 Multitouch Input Modality

The multi-touch input modality allows the user to perform the actions provided by the node array document model as described in detail in Chapter 5.2.1 by performing multi-touch gestures on the multi-touch screen or multi-touch pad of the mobile and wearable device. With multi-touch input pointers act relative to coordinates of the device.

This input modality is mainly purposed for smart-phones, smart-pablets and smart-tablets because these category of devices typically have a large multi-touch screen embedded and may be too large and heavy to be moved around themselves for using motion as with the orientation motion input modality proposed in the next Chapter 5.2.6.

According to these guidelines, the structure axis is mapped to the vertical y-axis of the multi-touch screen or multi-touch pad of the mobile and wearable device from the top to the bottom in this research. The text axis is mapped to the horizontal x-axis of the multi-touch screen or multi-touch pad of the mobile and wearable device from left to right. The insertElement axis is mapped to the y-axis of the multi-touch screen or multi-touch pad of the mobile and wearable device from to top to the bottom as with the structure axis. The following 15 actions of the node array model had been mapped to the following multi-touch gestures as shown in Table 5.8. Alternatives were not considered.

Action	Multi-touch Gesture
Set Structure Cursor	Put down one pointer at a specific position.
Move Structure Cursor	Move the one pointer up and down at an arbitrary speed.
Unset Structure Cursor	Lift up the one pointer.
Set Text Cursor	Move the one pointer more than 50 pixels left or right.
Move Text Cursor	Move the one pointer left or right at an arbitrary speed.
Unset Text Cursor	Move the one pointer more than 50 pixels up or down.
Set InsertElement Cursor	Slide in an additional second pointer from the top.
Move InsertElement Cursor	Move this second pointer up and down at an arbitrary speed.
Unset InsertElement Cursor	Lift up this second pointer.
Set Select Modifier	Put down an additional second or third pointer at an arbitrary position.
Unset Select Modifier	Lift up this second pointer.
Set Active Modifier	Put down an additional third pointer at an arbitrary position.
Unset Active Modifier	Lift up this third pointer.
Set InsertText Modifier	Slide in an additional second pointer from the left.
Unset InsertText Modifier	Lift up this second pointer.

Table 5.8: Multi-touch Gestures to Node Array Model Actions Mapping

5.2.6 Orientation Motion Input Modality

The orientation motion input modality allows the user to perform the actions provided by the node array document model as described in detail in Chapter 5.2.1 by performing motion gestures by moving the mobile or wearable device itself. With motion input, the device itself is acting relative to coordinates of the word. The orientation of a device is typically calculated by using a combination of the linear acceleration and magnetic field sensor, what is known as sensor fusion.

This input modality is mainly purposed for wearable devices like smart-watches because these category of devices typically are very small and lightweight and can therefore be moved around easily. In addition there may be only a very small multi-touch screen or multi-touch pad or no multi-touch screen or multi-touch pad at all embedded in these devices making the use of multi-touch input as with the multi-touch input modality proposed in the previous Chapter 5.2.5 difficult.

According to these guidelines, the structure axis is mapped to the horizontal orientation axis of the device anticlockwise within an angle of 180 degrees from the initial position of the device where the user interface was started in this research. The text axis is mapped to the vertical orientation axis of the device anticlockwise within an angle of 90 degrees starting at 0 degrees. The insertElement axis, as with the structure axis, is mapped to the horizontal orientation axis of the device anticlockwise within an angle of 180 degrees from the initial position of the device where the user interface was started. The following 15 actions of the node array model had been mapped to the following motion gestures as shown in Table 5.9. Alternatives were not considered.

Action	Motion Gesture
Set Structure Cursor	Rotate the device horizontally anticlockwise 10 degrees.
Move Structure Cursor	Rotate the device horizontally clockwise or anticlockwise within an angle of 180 degrees.
Unset Structure Cursor	Rotate the device horizontally back to the start position.
Set Text Cursor	Rotate the device vertically anticlockwise 10 degrees.
Move Text Cursor	Rotate the device vertically clockwise or anticlockwise within an angle of 90 degrees.
Unset Text Cursor	Rotate the device vertically back to the start position.
Set InsertElement Cursor	Flip the device 90 degrees to the right and rotate the device horizontally anticlockwise 10 degrees.
Move InsertElement Cursor	Rotate the device horizontally clockwise or anticlockwise within an angle of 180 degrees.
Unset InsertElement Cursor	Flip back the device 90 degrees to the left.
Set Select Modifier	Flip the device 90 degrees to the left.
Unset Select Modifier	Flip back the device 90 degrees to the right.
Set Active Modifier	Flip the device 180 degrees to the left.
Unset Active Modifier	Flip back the device 180 degrees to the right.
Set InsertText Modifier	Flip the device 90 degrees to the right and rotate the device horizontally anticlockwise 10 degrees.
Unset InsertText Modifier	Flip back the device 90 degrees to the left.

Table 5.9: Motion Gestures to Node Array Model Actions Mapping

5.3 User Interaction Design

The user interaction design describes the interactions on how the user acts in order to perform a specific task and how the user interface reacts on these actions performed by the user. The interactions are so designed that for each action performed by the user by using one of the different input modalities there is an immediate reaction of the user interface by all of the different output modalities. Figure 5.3 gives an overview over the user interaction design employed in this research:

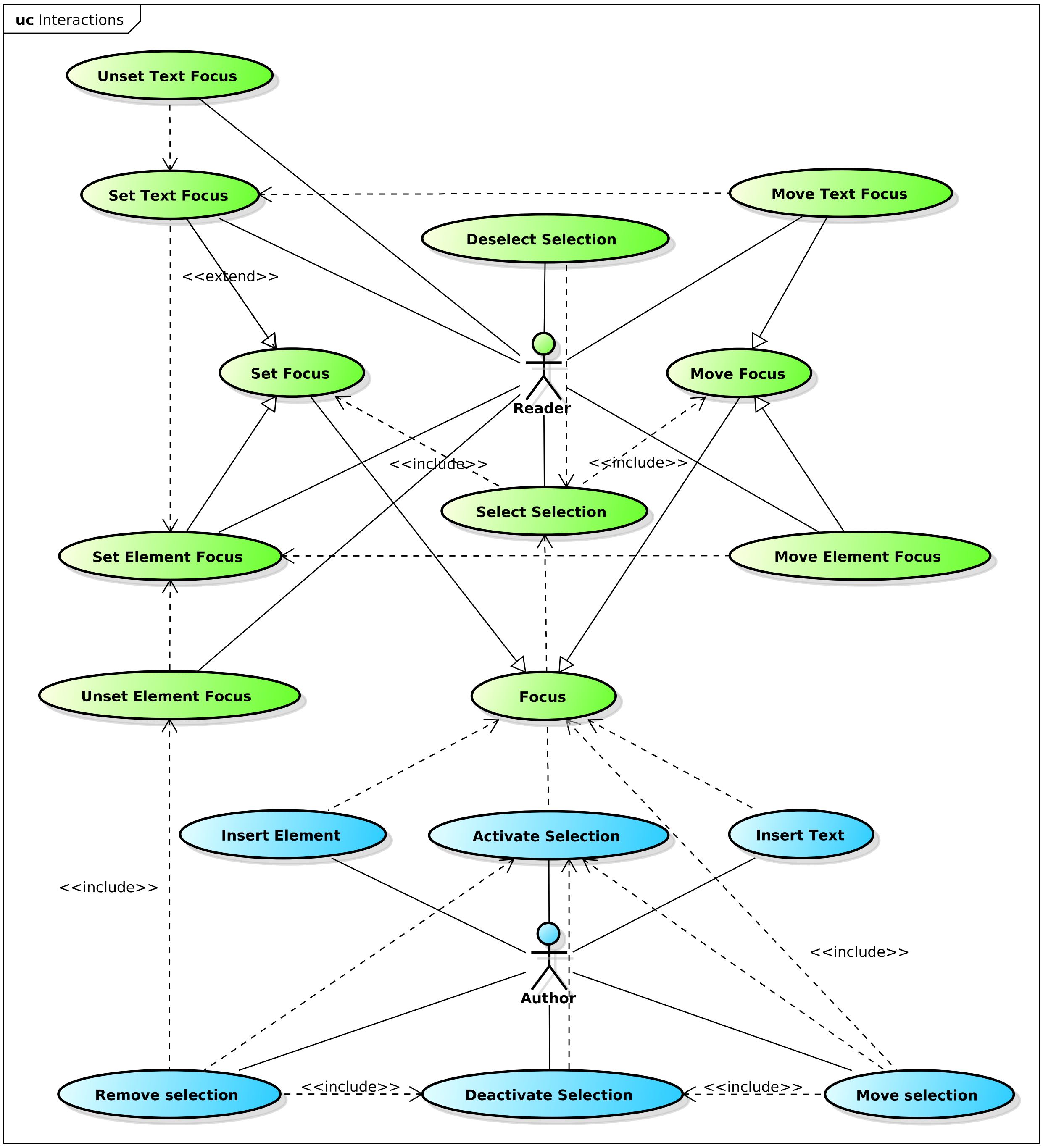


Figure 5.3: User Interaction Design Overview

There are the following two stakeholders with different interests. The reader wants to read a structured document by performing interactions for navigation like setting, moving and unsetting the element and text focus as well as to selecting a selection of the document containing one or more structure elements and / or texts. The author, on the other hand, is a more specialised type of reader, who wants to manipulate the structured document in addition to reading it by performing interactions for activate, deactivate, move, remove and insert structure elements and texts.

5.3.1 Set Element Focus

The reader wants to set the focus to a structure element within the document. He or she sets the structure cursor to a position on the structure axis by using one of the different input modalities. The user interface starts presenting an interaction with the operation focus for each structure element or text on each level at the current position of the document with all output modalities. The structure element on the topmost level receives the focus for further operations.

Stakeholders and Interests

- Reader: Wants to set the focus to a structure element within the document.

Pre Conditions

None

Main Success Scenario

1. The reader sets the structure cursor to a position on the structure axis by using one of the different input modalities.
2. The user interface starts presenting an interaction with the operation focus for each structure element or text on each level at the current position of the document with all output modalities. The structure element on the topmost level receives the focus for further operations.

Post Conditions

- A structure element within the document is focused for further operations.

5.3.2 Set Text Focus

The reader wants to set the focus to a word within the text at the current position of the document. He or she sets the text cursor to a position on the text axis by using one of the different input modalities. The user interface aborts presenting the focus interaction for the entire text at the current position within the document and starts a new interaction with the operation focus for the word at the current position within that text with all output modalities. This word receives the focus for further operations.

Stakeholders and Interests

- Reader: Wants to set the focus to a word within the text at the current position of the document.

Pre Conditions

- A structure element within the document is focused as described in detail in Chapter 5.3.1, Set Element Focus.

Main Success Scenario

1. The reader sets the text cursor to a position on the text axis by using one of the different input modalities.
2. The user interface aborts presenting the focus interaction for the entire text at the current position within the document and starts a new interaction with the operation focus for the word at the current position within that text with all output modalities. This word receives the focus for further operations.

Post Conditions

- A word within the text at the current position of the document is focused for further operations.

5.3.3 Move Element Focus

The reader wants to move the focus from the structure element which currently has the focus to another structure element within the document. He or she moves the structure cursor from the current position forward or backward to a new position on the structure axis by using one of the different input modalities. The user interface aborts presenting the focus interaction on that levels where the current structure element or text has been left and starts presenting new interactions with the operation focus on that levels where a new structure element or text has been entered with all output modalities. The structure element on the topmost level receives the focus for further operations.

Stakeholders and Interests

- Reader: Wants to move the focus from the structure element which currently has the focus to another structure element within the document.

Pre Conditions

- The focus is set to a structure element within the document as described in detail in Chapter 5.3.1, Set Element Focus.

Main Success Scenario

1. The reader moves the structure cursor from the current position to a new position on the structure axis by using one of the different input modalities.
2. The user interface aborts presenting the focus interaction on that levels where the current structure element or text has been left and starts presenting new interactions with the operation focus on that levels where a new structure element or text has been entered with all output modalities. The structure element on the topmost level receives the focus for further operations.

Post Conditions

- A new structure element within the document is focused for further operations.

5.3.4 Move Text Focus

The reader wants to move the focus from the word which currently has the focus to another word within the text at the current position of the document. He or she moves the text cursor from the current position forward or backward to a new position on the text axis by using one of the different input modalities. The user interface aborts presenting the focus interaction for the current word and starts presenting a new interaction with the operation focus for the word at the new position within the text at the current position of the document with all output modalities. This word receives the focus for further operations.

Stakeholders and Interests

- Reader: Wants to move the focus from the word which currently has the focus to another word within the text at the current position of the document.

Pre Conditions

- The focus is set to a word within the text at the current position of the document as described in detail in Chapter 5.3.2, Set Text Focus.

Main Success Scenario

1. The reader moves the text cursor from the current position to a new position on the text axis by using one of the input modalities.
2. The user interface aborts presenting the focus interaction for the current word and starts presenting a new interaction with the operation focus for the word at the new position within the text at the current position of the document with all output modalities. This word receives the focus for further operations.

Post Conditions

- A new word within the text at the current position of the document is focused for further operations.

5.3.5 Unset Text Focus

The reader wants to unset the focus from the word which currently has the focus within the text at the current position of the document. He or she unsets the text cursor at the current position on the text axis by using one of the different input modalities. The user interface aborts presenting the focus interaction for the current word and starts presenting a new interaction with the operation focus for the entire text at the current position within the document with all output modalities. The structure element on the topmost level, in which this text is contained, receives the focus for further operations. No word within this text is focused.

Stakeholders and Interests

- Reader: Wants to unset the focus from the word which currently has the focus within the text at the current position of the document.

Pre Conditions

- The focus is set to a word within the text at the current position of the document as described in detail in Chapter 5.3.2, Set Text Focus.

Main Success Scenario

1. The reader unsets the text cursor at the current position on the text axis by using one of the different input modalities.
2. The user interface aborts presenting the focus interaction for the current word and starts presenting a new interaction with the operation focus for the entire text at the current position within the document with all output modalities. The structure element on the topmost level, in which this text is contained, receives the focus for further operations. No word within this text is focused.

Post Conditions

- No word within the text at the current position of the document is focused.

5.3.6 Unset Element Focus

The reader wants to unset the focus from the structure element which currently has the focus within the document. He or she unsets the structure cursor at the current position on the structure axis by using one of the different input modalities. The user interface aborts presenting the focus interaction for all structure elements or texts on all levels at the current position within the document with all output modalities. No structure element is focussed for further operations.

Stakeholders and Interests

- Reader: Wants to unset the focus from the structure element which currently has the focus within the document.

Pre Conditions

- The focus is set to a structure element within the document as described in detail in Chapter 5.3.1, Set Element Focus.

Main Success Scenario

1. The reader unsets the structure cursor at the current position on the structure axis by using one of the different input modalities.
2. The user interface aborts presenting the focus interaction for all structure elements and texts on all levels at the current position of the document with all output modalities. No structure element is focussed for further operations.

Post Conditions

- No structure element within the document is focused.

5.3.7 Select Selection

The reader wants to select a selection of the document containing one or more structure elements and / or texts. He or she sets the select modifier by using one of the different input modalities and sets the focus to the start position of the current portion of the selection within the document. The user interface starts presenting a new interaction with the operation select for the structure element or text which currently has the focus with all output modalities. This structure element or text is selected. The reader optionally moves the focus to the end position of the current portion of the selection within the document. The user interface starts presenting a new interaction with the operation select for all structure elements and texts which are between the start and the end position of the current portion of the selection with all output modalities. These structure elements and text are included in the selection.

Stakeholders and Interests

- Reader: Wants to select a selection of the document containing one or more structure elements and / or texts.

Pre Conditions

None

Main Success Scenario

1. The reader sets the select modifier by using one of the different input modalities.
2. The reader sets the focus to the start position of the current portion of the selection within the document as described in detail in Chapter 5.3.1, Set Element Focus and Chapter 5.3.2, Set Text Focus.
3. The user interface starts presenting a new interaction with the operation select for the structure element or text which currently has the focus with all output modalities. This structure element or text is selected.
4. (Optional) The reader moves the focus to the end position of the current portion of the selection within the document as described in detail in Chapter 5.3.3, Move Element Focus and Chapter 5.3.4, Move Text Focus.

5. (Optional) The user interface starts presenting a new interaction with the operation select for all structure elements and texts which are between the start and the end position of the current portion of the selection with all output modalities. These structure elements and text are included in the selection.
6. (Optional) The reader repeats step 3 to 5 in order to add more portions to the selection.

Post Conditions

- A selection of the document containing one or more structure elements and / or texts is selected.

5.3.8 Deselect Selection

The reader wants to deselect the currently selected selection of the document. He or she unsets the select modifier by using one of the different input modalities. The user interface aborts presenting the select interactions for all structure elements and texts contained in the selection with all output modalities. No selection of the document is selected.

Stakeholders and Interests

- Reader: Wants to deselect the currently selected selection of the document.

Pre Conditions

- A selection of the document is selected as described in detail in Chapter 5.3.7, Select Selection.

Main Success Scenario

1. The reader unsets the select modifier by using one of the different input modalities.
2. The user interface aborts presenting the select interactions for all structure elements and texts contained in the selection with all output modalities. No selection of the document is selected.

Post Conditions

- No selection of the document is selected.

5.3.9 Activate Selection

The author wants to activate the currently selected selection of the document for manipulation. He or she sets the active modifier by using one of the different input modalities. The user interface starts presenting new interactions with the operation active for all structure elements and texts contained in the selection with all output modalities. The currently selected selection of the document is activated for manipulation.

Stakeholders and Interests

- Author: Wants to activate the currently selected selection of the document for manipulation.

Pre Conditions

- A selection of the document is selected as described in detail in Chapter 5.3.7, Select Selection.

Main Success Scenario

1. The author sets the active modifier by using one of the different input modalities.

2. The user interface starts presenting new interactions with the operation active for all structure elements and texts contained in the selection with all output modalities. The currently selected selection of the document is activated for manipulation.

Post Conditions

- The currently selected selection of the document is activated for manipulation.

5.3.10 Deactivate Selection

The author wants to deactivate the currently selected selection of the document for manipulation. He or she unsets the active modifier by using one of the different input modalities. The user interface aborts presenting the active interactions for all structure elements and texts contained in the selection with all output modalities. The currently selected selection of the document is deactivated for manipulation.

Stakeholders and Interests

- Author: Wants to deactivate the currently selected selection of the document for manipulation.

Pre Conditions

- The currently selected selection of the document is activated for manipulation as described in detail in Chapter 5.3.9, Activate Selection.

Main Success Scenario

1. The author unsets the active modifier by using one of the different input modalities.
2. The user interface aborts presenting the active interactions for all structure elements and texts contained in the selection with all output modalities. The currently selected selection of the document is deactivated for manipulation.

Post Conditions

- The currently selected selection of the document is deactivated for manipulation.

5.3.11 Move Selection

The author wants to move the currently selected and activated selection of the document to another position within the document. He or she focuses the target position within the document where the selection should be moved to. The user interface aborts presenting the active interactions and starts presenting a new interaction with the operation move for each structure element and text contained in the selection with all output modalities. The author unsets the active modifier by using one of the different input modalities. The user interface completes presenting the move interaction for each structure element and text contained in the selection with all output modalities. The selection is moved to the target position within the document.

Stakeholders and Interests

- Author: Wants to move the currently selected and activated selection of the document to another position within the document.

Pre Conditions

- A selection of the document is selected and activated for manipulation as described in detail in Chapter 5.3.9, Activate Selection.

Main Success Scenario

1. The author focuses the target position within the document to where the selection should be moved as described in detail in Chapter 5.3.1, Set Element Focus to Chapter 5.3.4, Move Text Focus.
2. The user interface aborts presenting the active interactions and starts presenting a new interaction with the operation move for each structure element and text contained in the selection with all output modalities.
3. The author unsets the active modifier by using one of the different input modalities.
4. The user interface completes presenting the move interaction for each structure element and text contained in the selection with all output modalities. The selection is moved to the target position within the document.

Alternative Flows

3a. The author focuses back the current position of the selection as described in detail in Chapter 5.3.1, Set Element Focus to Chapter 5.3.4, Move Text Focus.

4a. The user interface aborts presenting the move interactions and starts presenting a new interaction with the operation active for each structure element and text contained in the selection with all output modalities. The process of moving the selection to another position within the document is aborted.

Post Conditions

- The currently selected and activated selection of the document is moved to another position within the document.

5.3.12 Remove Selection

The author wants to remove the currently selected and activated selection from the document. He or she unsets the focus from the structure element which currently has the focus. The user interface aborts presenting the active interactions and starts presenting a new interaction with the operation remove for each structure element and text contained in the selection with all output modalities. The author unsets the active modifier by using one of the different output modalities. The user interface completes presenting the remove interaction for each structure element and text contained in the selection with all output modalities. The selection is removed from the document.

Stakeholders and Interests

- Author: Wants to remove the currently selected and activated selection from the document.

Pre Conditions

- A selection of the document is selected and activated for manipulation as described in detail in Chapter 5.3.9, Activate Selection.

Main Success Scenario

1. The author unsets the focus from the structure element which currently has the focus as described in detail in Chapter 5.3.6, Unset Element Focus.
2. The user interface aborts presenting the active interactions and starts presenting a new interaction with the operation remove for each structure element and text contained in the selection with all output modalities.

3. The author unsets the active modifier by using one of the different output modalities.
4. The user interface completes presenting the remove interaction for each structure element and text contained in the selection with all output modalities. The selection is removed from the document.

Alternative Flows

3a. The author sets the focus back to a structure element or text contained in the selection as described in detail in Chapter 5.3.1, Set Element Focus and Chapter 5.3.2, Set Text Focus.

4a. The user interface aborts presenting the remove interactions and starts presenting a new interaction with the operation active for each structure element and text contained in the selection with all output modalities. The process of removing the selection from the document is aborted.

Post Conditions

- The currently selected and activated selection of the document is removed from the document.

5.3.13 Insert Element

The author wants to insert a new structure element into the document after the currently focused position. He or she sets the insertElement cursor to a position on the insertElement axis by using one of the different input modalities. The user interface starts presenting a new interaction with the operation insert and the element type at the current position on the insertElement axis for the new structure element to be inserted. The author optionally moves the insertElement cursor on the insertElement axis forward or backward to choose another element type. The user interface aborts presenting the insert interaction with the current element type and starts presenting a new interaction with the operation insert and the new element type for the new structure element to be inserted. The author sets the select modifier by using one of the different input modalities. The user interface completes presenting the insert interaction for the new structure element to be inserted. A new structure element with that element type is created and inserted into the document after the currently focused position.

Stakeholders and Interests

- Author: Wants to insert a new structure element into the document after the currently focused position.

Pre Conditions

- A position within the document after which the new structure element should be inserted is focused as described in detail in Chapter 5.3.1, Set Element Focus to Chapter 5.3.4, Move Text Focus.

Main Success Scenario

1. The author sets the insertElement cursor to a position on the insertElement axis by using one of the different input modalities.
2. The user interface starts presenting a new interaction with the operation insert and the element type at the current position on the insertElement axis for the new structure element to be inserted.

3. (Optional) The author moves the insertElement cursor on the insertElement axis forward or backward to choose another element type.
4. (Optional) The user interface aborts presenting the insert interaction with the current element type and starts presenting a new interaction with the operation insert and the new element type for the new structure element to be inserted.
5. The author sets the select modifier by using one of the different input modalities.
6. The user interface completes presenting the insert interaction for the new structure element to be inserted. A new structure element with that element type is created and inserted into the document after the currently focused position.

Alternative Flows

- 5a. The author unsets the insertElement cursor at the current position on the insertElement axis by using one of the different input modalities.
- 6a. The user interface aborts presenting the insert interaction for the new structure element to be inserted. The process of inserting a new structure element into the document is aborted.

Post Conditions

- A new structure element is inserted into the document after the currently focused position.

5.3.14 Insert Text

The author wants to insert a new text into the document after the currently focused position. He or she sets the insertText modifier by using one of the different input modalities. The user interface starts presenting a new interaction with the operation insert for the new text to be inserted. The author speaks the text he or she wants to insert and unsets the insertText modifier by using one of the different input modalities. The user interface completes presenting the insert interaction for the new text to be inserted. A new text is created and inserted into the document after the currently focused position.

Stakeholders and Interests

- Author: Wants to insert a new text into the document after the currently focused position.

Pre Conditions

- A position within the document after which the new text should be inserted is focused as described in detail in Chapter 5.3.1, Set Element Focus to Chapter 5.3.4, Move Text Focus.

Main Success Scenario

1. The author sets the insertText modifier by using one of the different input modalities.
2. The user interface start presenting a new interaction with the operation insert for the new text to be inserted.
3. The author speaks the text he or she wants to insert.
4. The author unsets the insertText modifier by using one of the different input modalities.

5. The user interface completes presenting the insert interaction for the new text to be inserted. A new text is created and inserted into the document after the currently focused position.

Post Conditions

- A new text is inserted into the document after the currently focused position.

5.4 Conclusions

A multi-modal user interface for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices has been developed as a result of the inductive research in the previous chapters.

The structured document is provided to the user interface through the W3C Document Object Model (DOM). This is a platform and language neutral application programming interface which allows the user interface to dynamically access and update the document structure and content. In the DOM, documents are represented as a hierarchy of node objects which is very much like a tree.

In the next step, the document is reorganised and transformed into the Node Array Document Model, which is an alternative modality-neutral document model for presentation, navigation and manipulation of a structured documents. In contrast to the tree-like document representation employed in the Document Object Model described above, the element and text objects of the document are organised in a two-dimensional array. The advantages of this representation is that it enables the user to get a fast overview over the document structure and to efficiently skim and scan over the document content by identifying the type, level, position, length, relationship and text of each element.

The earcon output modality presents the interactions of the node array document model using earcons, which are non-verbal audio messages that are used in the computer user interface to provide information to the user about some computer object, operation or interaction. Each interaction is presented by a single-pitch inherited elementary earcon inheriting the timbre of the element type, pitch of the element level, spatial location of the element position, register of the operation and dynamics of the status. These single-pitch inherited elementary earcons are rendered simultaneously as parallel compound earcons for all interactions that are going on at the same time to reduce the length of time a compound audio message takes.

The tacton output modality presents the interactions of the node array document model using tactons or tactile icons. These are structured, abstract vibrotactile messages that can be used for presenting multidimensional information non-visually. Each interaction is presented by a single-motive inherited elementary tacton inheriting the rhythm of the element type, the tempo of the element level and the intensity of the operation. These single-motive inherited elementary tactons are rendered sequentially as serial compound tactons for all interactions that are going on at the same time.

The speech output modality presents the interactions of the node array document model using synthetic speech. Each interaction is presented by a inherited elementary utterance, inheriting the voice of the element type, pitch of the node level, spatial location of the node position, register of the operation and text of the content. These inherited elementary utterances are rendered sequentially as serial compound utterances for all interactions that are going on at the same time.

The multi-touch input modality allows the user to perform the actions provided by the node array document model by performing multi-touch gestures on the multi-touch screen or multi-touch pad of the mobile and wearable device. This input modality is mainly purposed for smart-phones, smart-pablets and smart-tablets because these category of devices typically have a large multi-touch screen embedded and may be too large and heavy to be moved around themselves for using motion.

The orientation motion input modality allows the user to perform the actions provided by the node array document model by performing motion gestures by moving the mobile or wearable device itself. This input modality is mainly purposed for wearable devices like smart-watches because these category of devices typically are very small and lightweight and can therefore be moved around easily. In addition there may be only a very small multi-touch screen or multi-touch pad or no multi-touch screen or multi-touch pad at all embedded in these devices making the use of multi-touch input difficult.

There are two stakeholders with different interests. The reader wants to read a structured document by performing interactions for navigation like setting, moving and unsetting the element and text focus as well as to selecting a selection of the document containing one or more structure elements and / or texts. The author, on the other hand, is a more specialised type of reader, who wants to manipulate the structured document in addition to reading it by performing interactions for activate, deactivate, move, remove and insert structure elements and text. These interactions are so designed that for each action performed by the user by using one of the different input modalities there is an immediate reaction of the user interface by all of the different output modalities.

6 Investigation into the Effectiveness of the DOKY User Interface by Automated Structured Observation of Blind and Visually Impaired Persons

6.1 Introduction

This chapter presents the results of an investigation into the effectiveness of the DOKY user interface proposed in the previous Chapter 5, which was carried out to see whether the proposed user interface design concepts and user interaction design concepts are effective means for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices, by automated structured observations of 876 blind and visually impaired research subjects performing 19 exercises among a highly structured example document using the DOKY Structured Observation App as introduced in Appendix C on their own mobile or wearable device remotely over the Internet. If they are not then there would be no point in using them and alternative methods would have to be found. The research methodology employed has been introduced in detail in Chapter 3.3.

An important point to keep in mind is that no training has been provided to the research subjects prior to this structured observation. The work of Brewster, Wright & Edwards (1993) indicated that the more earcons were heard the better the recognition rates would be. This indicated that, if earcons were used at the human computer user interface, then regular exposure would quickly lead to high levels of recognition. Their results also indicated that musicians have again been shown to be no better than non-musicians. This means that auditory interfaces will be usable by most users whatever their level of musical skill. The same may also be true for to the tactons, speech utterances, multitouch gestures and motion gestures employed in the DOKY user interface.

6.2 Research Subjects

Different organisations in support of blind and visually impaired persons in the United Kingdom, the United States, Germany, Austria and Switzerland had been selected and invited to participate in this research for sampling the blind and visually impaired research subjects for this structured observation as described in detail in Chapter 3.2.4.

Out of this blind and visually impaired persons who had been invited by the different organisations to participate in this structured observation, feedbacks from 876 respondents, forming the sample for this research, had been received in different regions of the world as well as with different languages, using different operating systems on various different hardware devices, with and without screen reader. This sample is described in more detail in this chapter.

The response rate for this activity was notably over four times greater than with the initial survey because the selection has been extended to the United Kingdom and the United States and the research subjects were very interested and enthusiastic in testing a novel concept for an assistive technology in support of blind and visually impaired persons for the reading of structured documents rather than just answering questions.

6.2.1 Region

The research subjects taking part in this research are located in different regions of the world. In the United Kingdom, the United States, Germany, Austria and Switzerland, blind and visually impaired persons had been invited through organisations to take part in this research. In addition, research subjects came from other regions which had not been invited directly. These regions are briefly described below. Figure 6.1 shows the different regions of the world in which the persons concerned are located as well as the frequencies and percentages of respondents in each region:

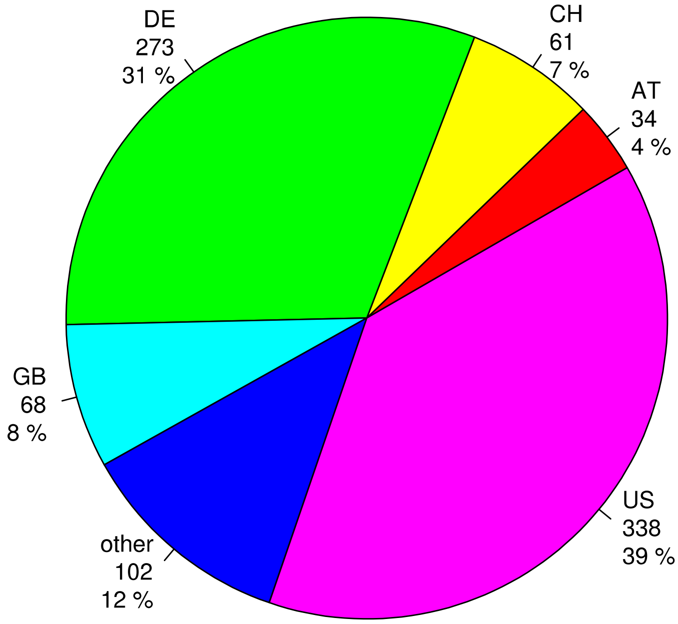


Figure 6.1: Regions of Research Subjects

39 per cent of the research subjects are in the United States, 31 per cent of them are in Germany, 8 per cent of the persons concerned came from the United Kingdom, 7 per cent of the respondents from Switzerland, 4 per cent of them from Austria and 12 per cent from other countries which had not been invited directly.

6.2.2 Language

The research subjects requested different languages for taking part in this research. The DOKY Structured Observation App supports the English and the German language. If a research subject requested another language, which is not supported, English is served as the default language. These languages are briefly described below. Figure 6.2 shows the different languages requested by the persons concerned as well as the frequencies and percentages of respondents requesting each language:

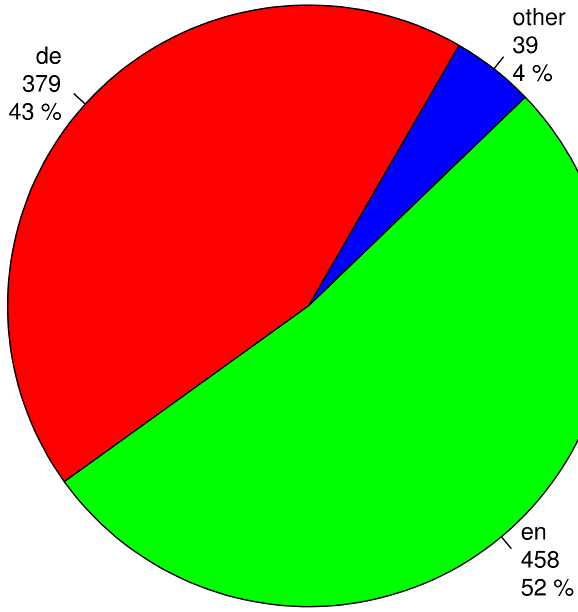


Figure 6.2: Languages requested by Research Subjects

52 per cent of the research subjects requested the English language, 43 per cent of them requested the German language and only 4 per cent of the persons concerned requested another language which is not supported. In this case English was served because it is the default language.

6.2.3 Operating System

The research subjects used different operating systems for taking part in this research. The DOKY Structured Observation App is provided for the Apple iOS and the Google Android operating system. These operating systems are briefly described below. Figure 6.3 shows the frequencies and percentages of respondents using each of these two operating systems:

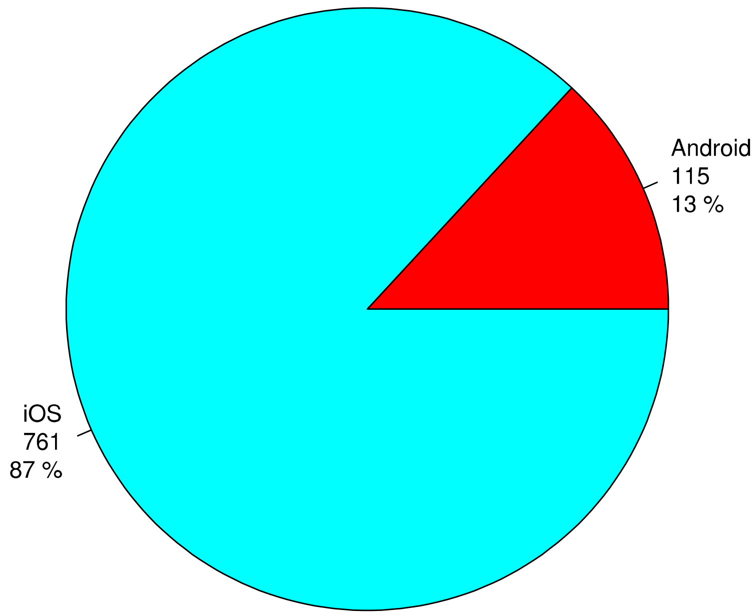


Figure 6.3: Operating Systems used by Research Subjects

Most of the blind and visually impaired research subjects, namely 87 per cent of them, used Apple iOS and they are accompanied by only 13 per cent of Google Android users because Apple's iDevices are very popular with blind and visually impaired persons due to the fact that Apple was the first manufacturer, providing assistive technologies like screen readers and screen magnifiers out of the box and free of charge.

6.2.4 Hardware Device

The research subjects employed different hardware devices for taking part in this research. These hardware devices are briefly described below. Figure 6.4 shows the different hardware devices used by the persons concerned as well as the frequencies and percentages of respondents using each hardware device:

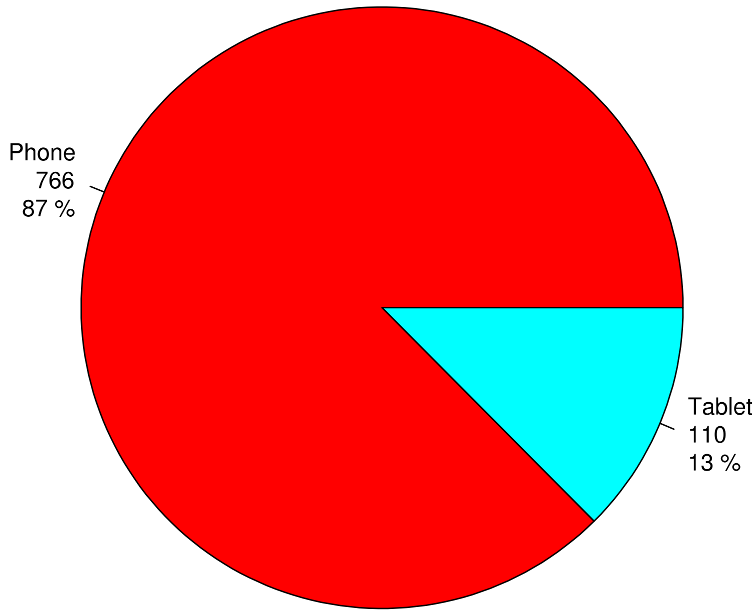


Figure 6.4: Hardware Devices used by Research Subjects

Most of the research subjects, namely 87 per cent of them, used a smart phone and only 13 per cent of the them used a smart tablet. Nobody of the persons concerned used a smart watch because this category of hardware device is very new and therefore not already employed by the respondents. For this reason, wearable devices were not tested and motion interaction was only tested to a limited degree.

6.2.5 Screen Reader

Some of the research subjects taking part in this research used a screen reader like Google [TalkBack] or Apple [VoiceOver] on their mobile or wearable device. Figure 6.5 shows the frequencies and percentages of respondents in each of the two categories of screen reader users:

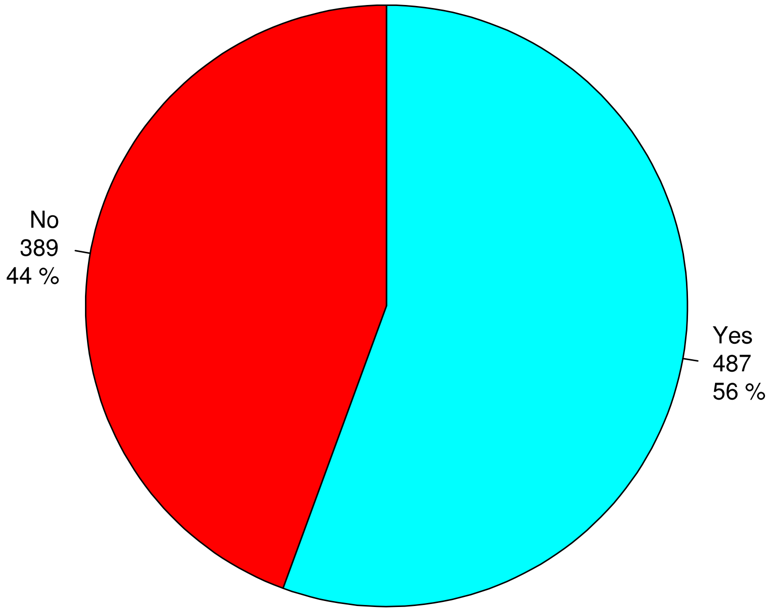


Figure 6.5: Use of Screen Reader by Research Subjects

56 per cent of the research subjects used a screen reader like Google [TalkBack] or Apple [VoiceOver] and 44 per cent of the persons concerned do not use a screen reader.

The use of screen readers depends to 5 per cent on the hardware device used, as shown by a weak bivariate correlation with v = 0.23 and a very high level of statistical significance with p = 2.41e-12. Figure 6.6 shows the use of screen readers in correlation to the different categories of hardware devices as well as the percentages of how often they are used by each hardware device category:

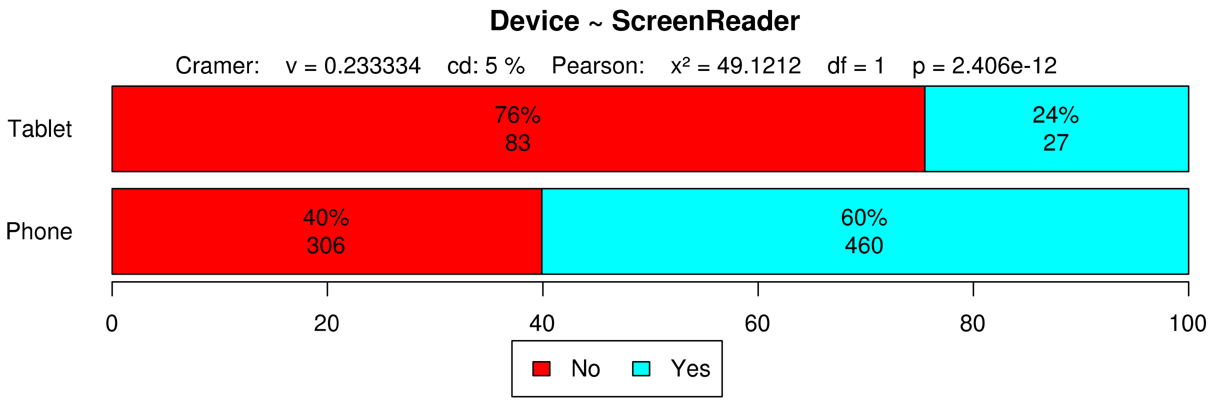


Figure 6.6: Use of screen readers in correlation to Hardware Device Category

60 per cent of the research subjects using a smart phone as well as 24 per cent of them using a smart tablet are also using a screen reader. On the other hand, 76 per cent of the respondents using a smart tablet and 40 per cent of them using a smart phone do not use screen readers. This correlation arises due to the fact that persons who are using screen readers and are not relying on the visual sense do not benefit from the larger screen embedded in smart tablets making them more large and heavy to be moved around.

6.3 Presentation

The first 6 exercises focussing on the user interface design concepts and their hypotheses about the behaviour of blind and visually impaired persons among structured documents in relation to the reactions for presentation of the element type, level, position, length, relationship and text provided by the different output modalities which are the Earcon Output Modality as described in Chapter 5.2.2, the Tacton Output Modality as described in Chapter 5.2.3 and the Speech Output Modality as described in Chapter 5.2.4 using earcons, tactons and synthetic speech.

6.3.1 Element Type

In this exercise, the participants had to find and select the document structure element of the type “graphic” in order to evaluate if and how well the respondents are able to recognise the element type of a document structure element. The element type is presented to the user by the timbre of the single-pitch inherited elementary earcon, the rhythm of the single-motive inherited elementary tacton and the voice of the inherited elementary speech utterance.

The research subjects used different input modalities to find and select the document structure element of the requested type. Figure 6.7 shows the different input modalities used by the persons concerned as well as the frequencies and percentages of respondents using each input modality:

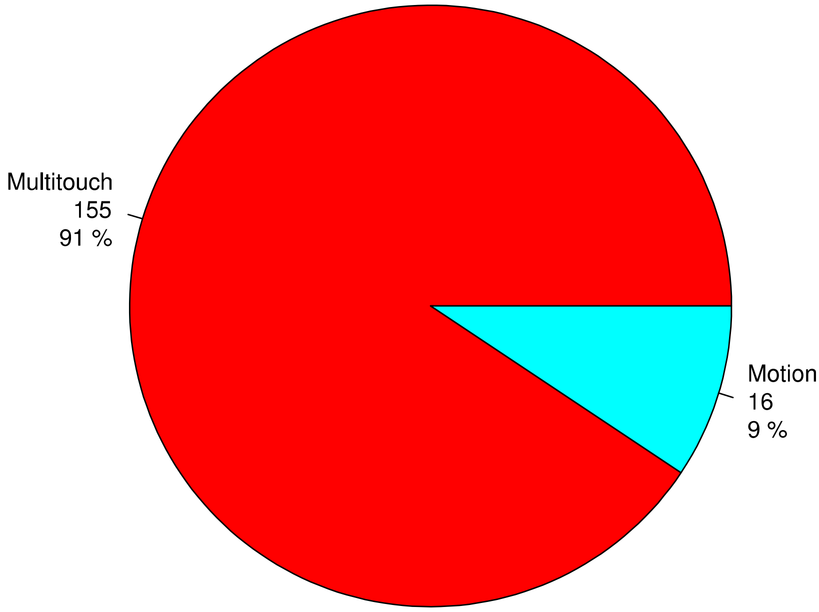


Figure 6.7: Input Modality used to Select the requested Element Type

Most of the research subjects, namely 91 per cent of them, used multitouch and only 9 per cent of the persons concerned used motion to find and select the document structure element of the requested type. Hypothesis 1, which proposes that blind and visually impaired persons who are using mobile devices like smart-phones and smart-tablets will choose to use the multitouch input modality to recognise the element type of a document structure element, can be confirmed.

The number of selects performed, time taken and number of interactions performed by the research subjects to find and select the document structure element of the requested type greatly differs between the cases. Figure 6.8 shows the central tendency, the median and dispersion as well as any outliers covering the entire number of selects, time and number of interactions range as well as frequencies and percentages of the detailed distribution of the interesting ranges only without the outliers:

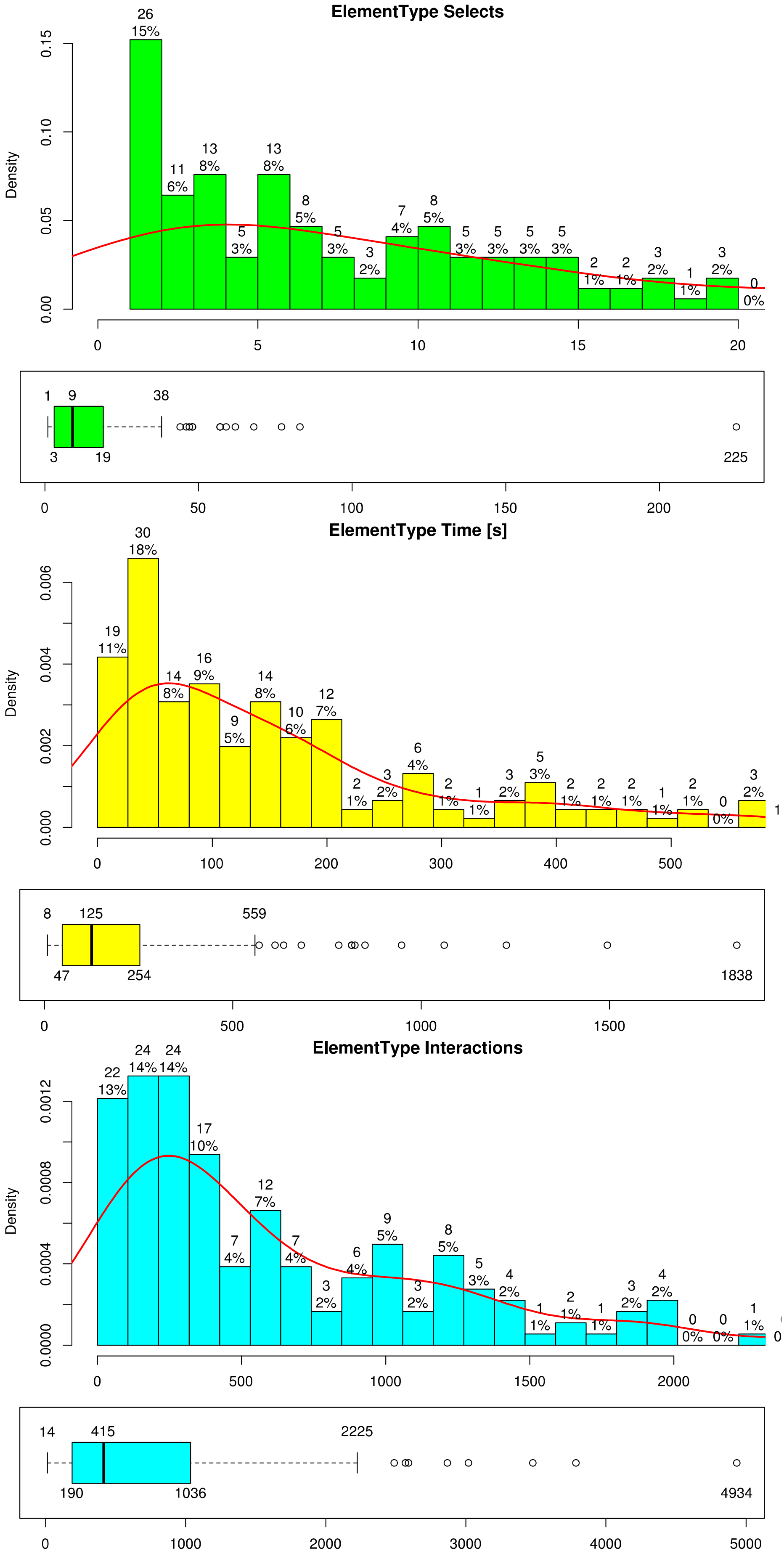


Figure 6.8: Selects, Time and Interactions to Select the requested Element Type

The middle 50 per cent of research subjects solved the task by performing 3 to 19 selects and the median by performing 9 selects. The most effective respondents solved the exercise by performing only 1 select and the participants less effective than any other, other than the outliers who might not participated seriously and performed up to 225 selects, with 38 selects. The greatest band of persons concerned within the detailed distribution, namely 15 per cent of them, solved the task by performing only 1 select.

The middle 50 per cent of research subjects solved the task within 47 to 254 seconds and the median in 125 seconds. The fastest respondents solved the exercise in only 8 seconds and the participants slower than any other, other than the outliers who might not participated seriously and took up to 1838 seconds, in 559 seconds. The greatest band of persons concerned within the detailed distribution, namely 18 per cent of them, solved the task within 25 and 50 seconds.

The middle 50 per cent of research subjects solved the task by performing 190 to 1036 interactions and the median by performing 415 interactions. The most effective respondents solved the exercise by performing only 14 interactions and the participants less effective than any other, other than the outliers who might not participated seriously and performed up to 4934 interactions, with 2225 interactions. The greatest band of persons concerned within the detailed distribution, namely 28 per cent of them, solved the task by performing between 100 and 300 interactions.

This confirms Hypothesis 2, which proposes that the user interface design concepts for presenting the element type will be effective and enable blind and visually impaired persons to recognise the element type of a document structure element with a recall rate which will be much better than the number of selects required to recognise the element type of a document structure element by chance only, within a usable amount of time taken and with a usable effort of performed interactions.

Hypothesis 3, which proposes that the same effectiveness (number of selects performed, time taken and number of interactions performed) will be achieved across all different input modalities, can be confirmed because no statistically significant correlation between the input modality used and the number of selects performed, time taken and number of interactions performed to find and select the document structure element of the requested type could be found.

The effectiveness is dependent on the use of screen reader by a specific person. The number of interactions performed depends to 2 per cent on the use of screen reader, as shown by a weak bivariate correlation with $v = 0.14$ and a high level of statistical significance with $p = 0.059$. Figure 6.9 shows the central tendency, the median and dispersion as well as any outliers covering the entire number of interactions range in correlation to the use of screen reader:

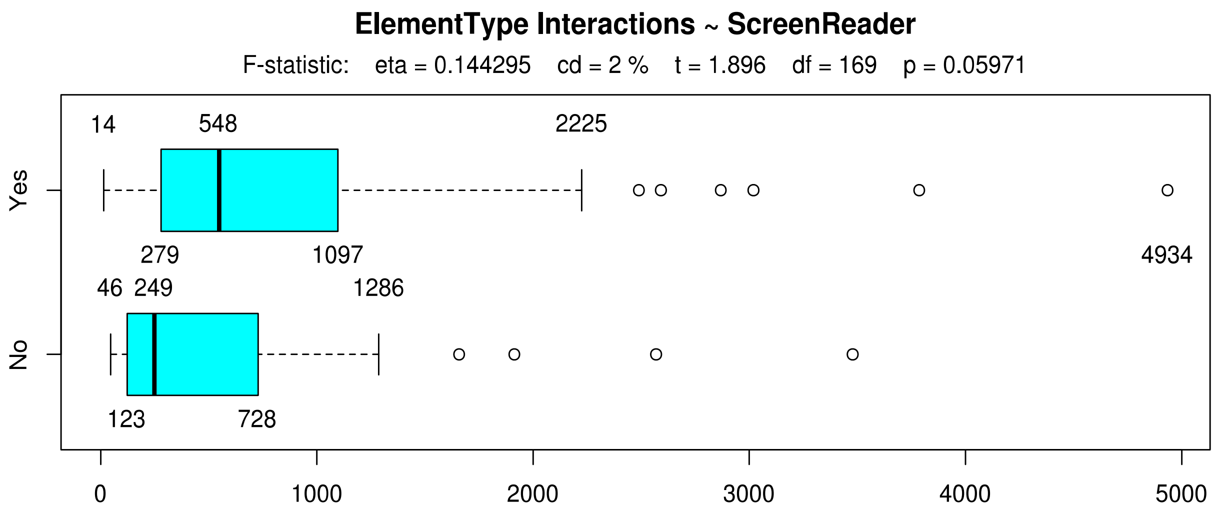


Figure 6.9: Element Type Interactions in correlation to use of Screen Reader

The middle 50 per cent of research subjects who are using screen reader solved the task by performing 279 to 1097 interactions and the median by performing 548 interactions. The most effective respondents solved the exercise by performing only 14 interactions and the participants less effective than any other, other than the outliers who performed up to 4934 interactions, with 2225 interactions. On the other hand, the middle 50 per cent of research subjects who are not using screen reader solved the task by performing 123 to 728 interactions and the median by performing 249 interactions. The most effective respondents solved the exercise by performing only 45 interactions and the participants less effective than any other, other than the outliers, with 1285 interactions.

Therefore Hypothesis 4, which proposes that blind and visually impaired persons who are using screen reader will perform better than respondents who are not using screen reader to find and select the document structure element of the requested type, can be rejected.

Hypothesis 5, which proposes that the same effectiveness will be achieved across all the different hardware devices employed, can be confirmed because no statistically significant correlation between the hardware device used and the number of selects performed, time taken and number of interactions performed to find and select the document structure element of the requested type could be found.

There is a multivariate relationship between the number of selects performed, the time taken and the number of interactions performed to solve the task of this exercise. The number of selects performed depends to 26 per cent on the time taken to solve the task, as shown by a moderate bivariate correlation with cor = 0.51 and a very high level of statistical significance with p = 1.43e-12 as well as to 33 percent on the number of interactions performed, as shown by another moderate bivariate correlation with cor = 0.57 and a very high statistical significance level of p = 4.44e-16. The time taken itself depends to 74 per cent on the number of interactions performed to solve the task, as shown by a very strong bivariate correlation with cor = 0.86 an a very high level of statistical significance with p = 0. Figure 6.10 shows the number of selects performed in correlation to the time taken and the number of interactions performed as well as the regression pane of this relationship:

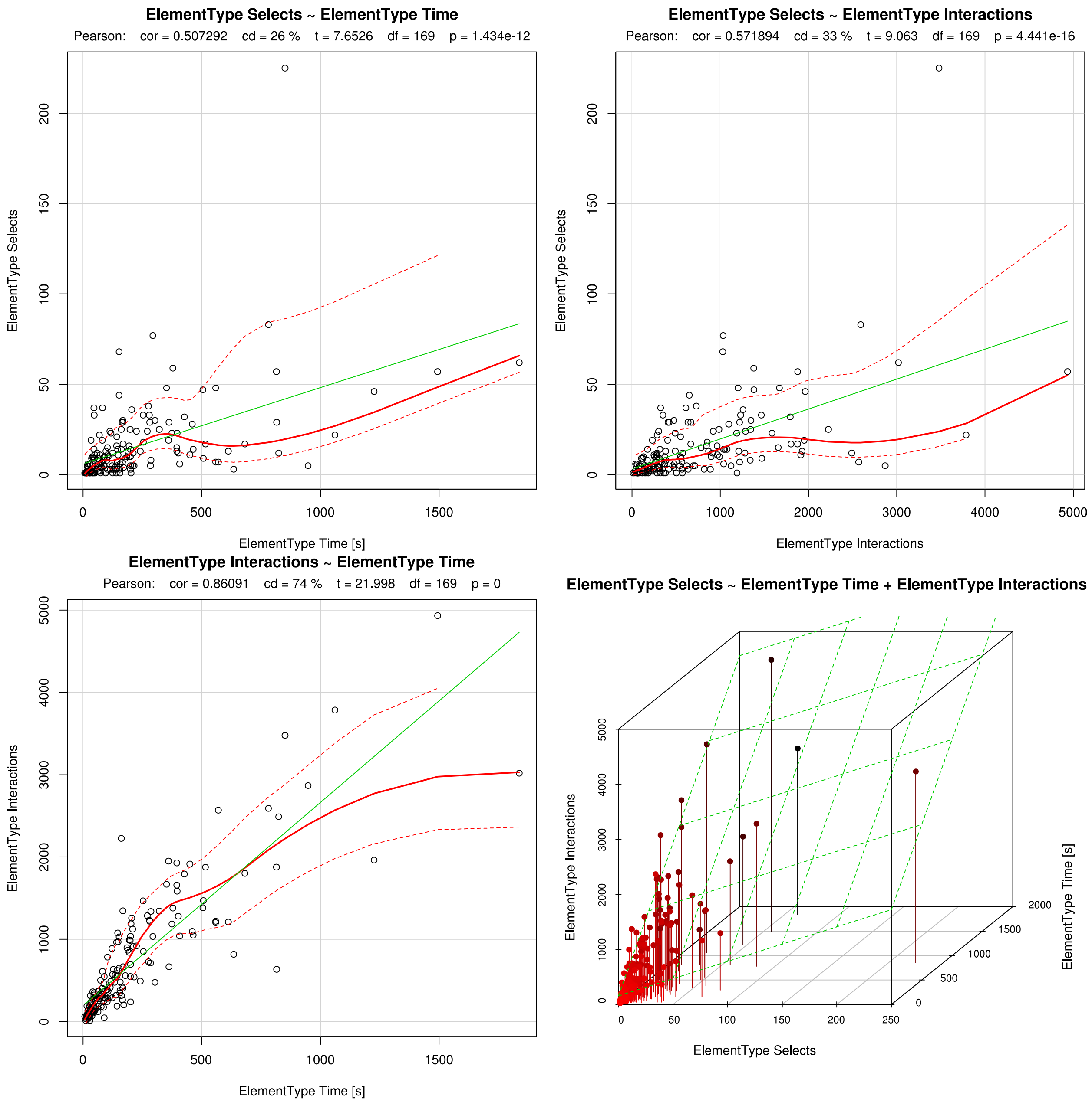


Figure 6.10: Element Type Selects in correlation to Time and Interactions

This correlation shows that the more selects had been performed, the more time was taken and the more interactions had been performed. These results confirm Hypothesis 6, which proposes that the more effective the user interface design concepts for presenting the element type are, the less selects will be performed, the less time will be taken and the less interactions will be performed.

6.3.2 Level

In this exercise, the participants had to find and select a document structured element at the 4th level in order to examine if and how well the respondents are able to recognise the level of a document structure element or text node. The level is presented to the user by the pitch of the single-pitch inherited elementary earcon, the tempo of the single-motive inherited elementary tacton and the pitch of the inherited elementary speech utterance.

The research subjects used different input modalities to find and select the document structure element at the requested level. Figure 6.11 shows the different input modalities used by the persons concerned as well as the frequencies and percentages of respondents using each input modality:

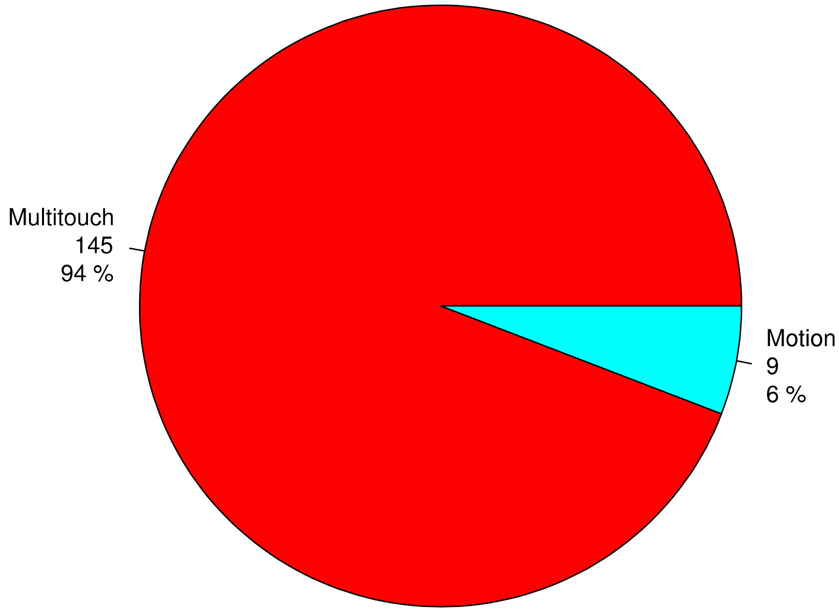


Figure 6.11: Input Modality used to Select the requested Level

Most of the research subjects, namely 94 per cent of them, used multitouch and only 6 per cent of the persons concerned used motion to find and select the document structure element at the requested level. Hypothesis 1, which proposes that blind and visually impaired persons who are using mobile devices like smart-phones and smart-tablets will choose to use the multitouch input modality to recognise the level of a document structure element or text node, can be confirmed.

The number of selects performed, time taken and number of interactions performed by the research subjects to find and select the document structure element at the requested level greatly differs between the cases. Figure 6.12 shows the central tendency, the median and dispersion as well as any outliers covering the entire number of selects, time and number of interactions range as well as frequencies and percentages of the detailed distribution of the interesting ranges only without the outliers:

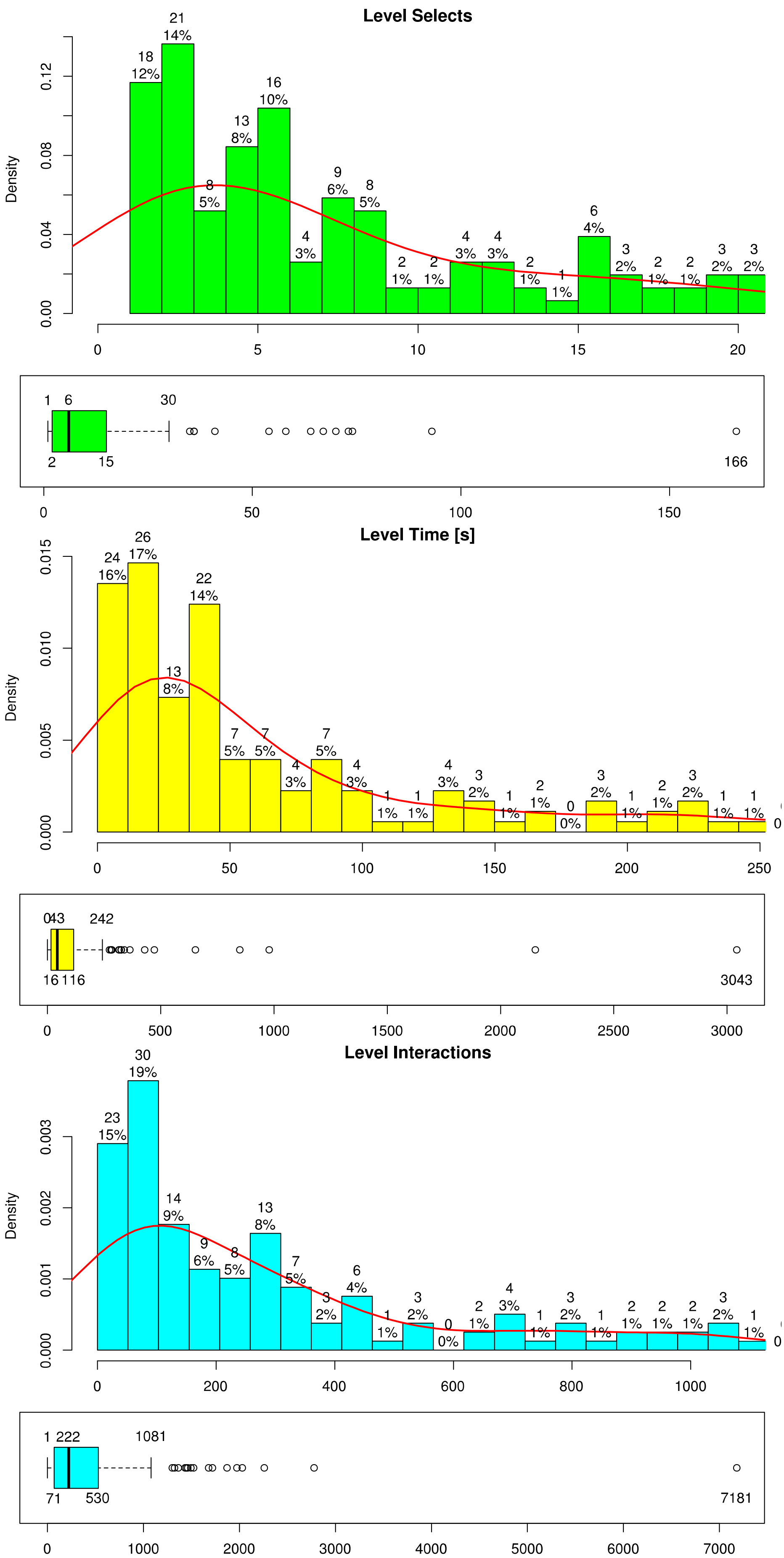


Figure 6.12: Selects, Time and Interactions to Select the requested Level

The middle 50 per cent of research subjects solved the task by performing 2 to 16 selects and the median by performing 6 selects. The most effective respondents solved the exercise by performing only 1 select and the participants less effective than any other, other than the outliers who might not participated seriously and performed up to 166 selects, with 30 selects. The greatest band of persons concerned within the detailed distribution, namely 14 per cent of them, solved the task by performing only 1 select.

The middle 50 per cent of research subjects solved the task within 16 to 116 seconds and the median in 43 seconds. The fastest respondents solved the exercise in less than 1 seconds and the participants slower than any other, other than the outliers who might not participated seriously and took up to 3043 seconds, in 242 seconds. The greatest band of persons concerned within the detailed distribution, namely 17 per cent of them, solved the task within 10 and 20 seconds.

The middle 50 per cent of research subjects solved the task by performing 190 to 1036 interactions and the median by performing 415 interactions. The most effective respondents solved the exercise by performing only 14 interactions and the participants less effective than any other, other than the outliers who might not participated seriously and performed up to 4934 interactions, with 2225 interactions. The greatest band of persons concerned within the detailed distribution, namely 28 per cent of them, solved the task by performing between 100 and 300 interactions.

This confirms Hypothesis 2, which proposes that the user interface design concepts for presenting the level will be effective and enable blind and visually impaired persons to recognise the level of a document structure element or text node with a recall rate which will be much better than the number of selects required to recognise the level of a document structure element or text node by chance only, within a usable amount of time taken and with a usable effort of performed interactions.

Hypothesis 3, which proposes that the same effectiveness (number of selects performed, time taken and number of interactions performed) will be achieved across all different input modalities, can be confirmed because no statistically significant correlation between the input modality used and the number of selects performed, time taken and number of interactions performed to find and select the document structure element at the requested level could be found.

Hypothesis 4, which proposes that blind and visually impaired persons who are using screen reader will perform better than respondents who are not using screen reader, can be rejected because no statistically significant correlation between the use of screen reader and the number of selects performed, time taken and number of interactions performed to find and select the document structure element at the requested level could be found.

The effectiveness is dependent on the hardware device used by a specific person. The number of selects performed depends to 6 per cent on the hardware device, as shown by a weak bivariate correlation with $v = 0.24$ and a very high level of statistical significance with $p = 0.0024$. Figure 6.13 shows the central tendency, the median and dispersion as well as any outliers covering the entire number of selects range in correlation to the different categories of hardware devices used:

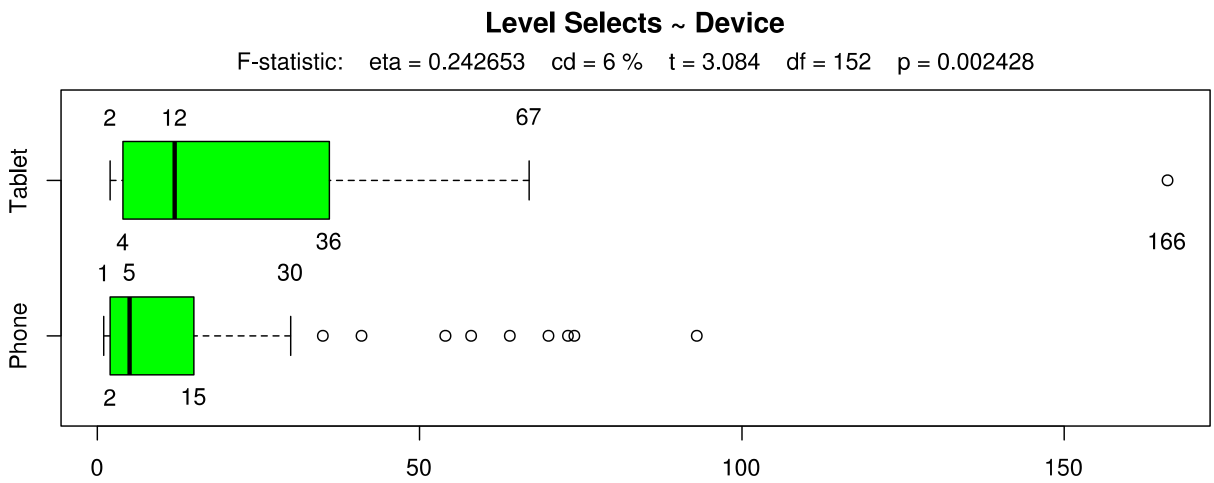


Figure 6.13: Level Selects in correlation to Hardware Device used

The middle 50 per cent of research subjects who solved the task by using a smart tablet performed 4 to 36 selects and the median performed 12 selects. The most effective respondents solved the exercise by performing only 2 selects and the participants less effective than any other, other than the outliers who performed up to 166 selects, with 67 selects. On the other hand, the middle 50 per cent of research subjects who solved the task by using a smart phone performed 2 to 16 selects and the median performed 5 selects. The most effective respondents solved the exercise by performing only 1 select and the participants less effective than any other, other than the outliers, with 30 selects.

This correlation arises due to the fact that Apple's smart tablets, and most of the blind and visually impaired research subjects, namely 87 per cent of them, used Apple devices as introduced in detail in Chapter 6.2.3, do not have a tactile stimulator built-in. Therefore the tacton output modality cannot be used and the persons concerned relied on the auditory human sense served by the earcon output modality and the speech output modality only making them less effective than the smart phone users who have all modalities available.

These results fit in well with the psychological evidence of Brown, Newsome & Glinert (1989), Perrott, Sadralobadi, Saberi & Strybel (1991) and Brewster (1992), who suggest that sharing information across different sensory modalities can actually improve task performance. Having redundant information gives the user multiple chances of identifying the data. For example, if they cannot remember how a document structure element level sounds like they may be able to remember what it feels like.

Therefore Hypothesis 5, which proposes that the same effectiveness will be achieved across all the different hardware devices employed to find and select the document structure element at the requested level, can be rejected.

There is a multivariate relationship between the number of selects performed, the time taken and the number of interactions performed to find and select the document structure element at the requested level. The number of selects performed depends to 23 per cent on the time taken to solve the task, as shown by a moderate bivariate correlation with cor = 0.48 and a very high level of statistical significance with p = 3.82e-10 as well as to 38 percent on the number of interactions performed, as shown by a strong bivariate correlation with cor = 0.61 and a very high statistical significance level of p = 0. The time taken itself depends to 51 per cent on the number of interactions performed to solve the task, as shown by a strong bivariate correlation with cor = 0.71 an a very high level of statistical significance with p = 0. Figure 6.14 shows the number of selects performed in correlation to the time taken and the number of interactions performed as well as the regression pane of this relationship:

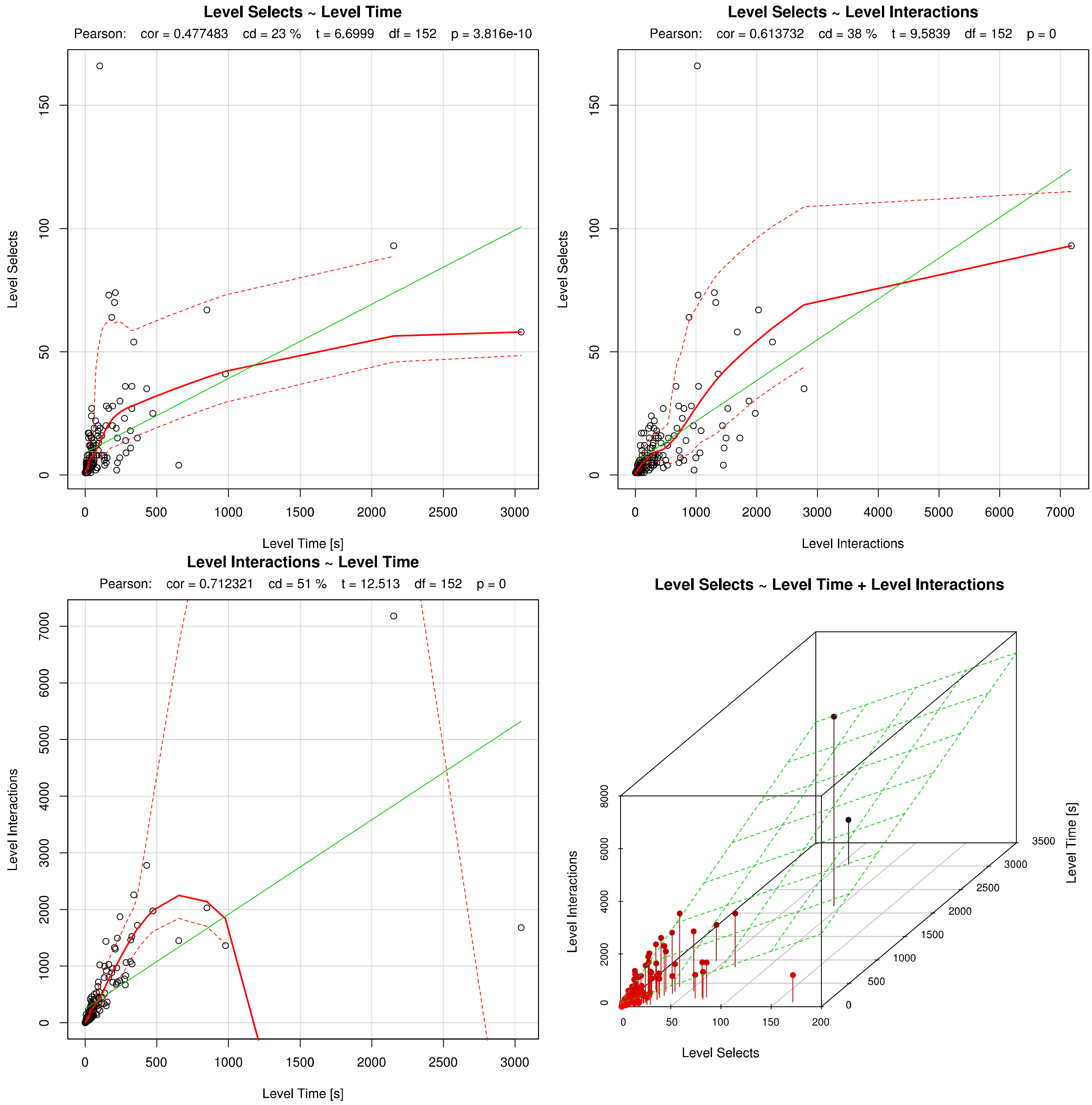


Figure 6.14: Level Selects in correlation to Time and Interactions

This correlation shows that the more selects had been performed, the more time was taken and the more interactions had been performed. These results confirm Hypothesis 6, which proposes that the more effective the user interface design concepts for presenting the level are, the less selects will be performed, the less time will be taken and the less interactions will be performed.

6.3.3 Position

In this exercise, the participants had to find and select the document structured element which is in the middle of the document in order to examine if and how well the respondents are able to recognise the position of a document structure element or text node. The position in document reading order is presented to the user by the spatial location of the single-pitch inherited elementary earcon, the position on the vertical y-axis of the multi-touch screen or multi-touch pad of the mobile and wearable device from the top to the bottom, the position on the horizontal orientation axis of the device anticlockwise within an angle of 180 degrees and the spatial location of the inherited elementary speech utterance.

The research subjects used different input modalities to find and select the document structure element at the requested position. Figure 6.15 shows the different input modalities used by the persons concerned as well as the frequencies and percentages of respondents using each input modality:

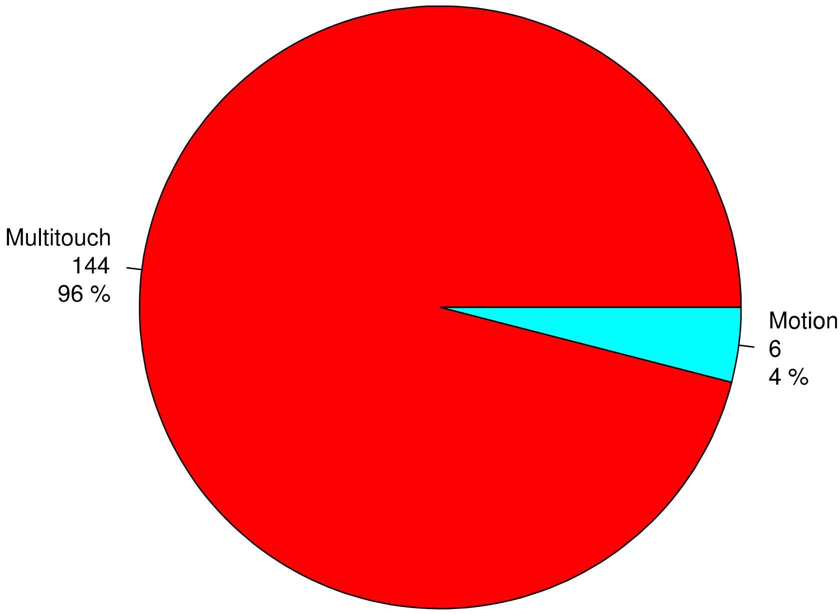


Figure 6.15: Input Modality used to Select the requested Position

Most of the research subjects, namely 96 per cent of them, used multitouch and only 4 per cent of the persons concerned used motion to find and select the document structure element at the requested position. Hypothesis 1, which proposes that blind and visually impaired persons who are using mobile devices like smart-phones and smart-tablets will choose to use the multitouch input modality to recognise the position of a document structure element or text node, can be confirmed.

The number of selects performed, time taken and number of interactions performed by the research subjects to find and select the document structure element at the requested position greatly differs between the cases. Figure 6.16 shows the central tendency, the median and dispersion as well as any outliers covering the entire number of selects, time and number of interactions range as well as frequencies and percentages of the detailed distribution of the interesting ranges only without the outliers:

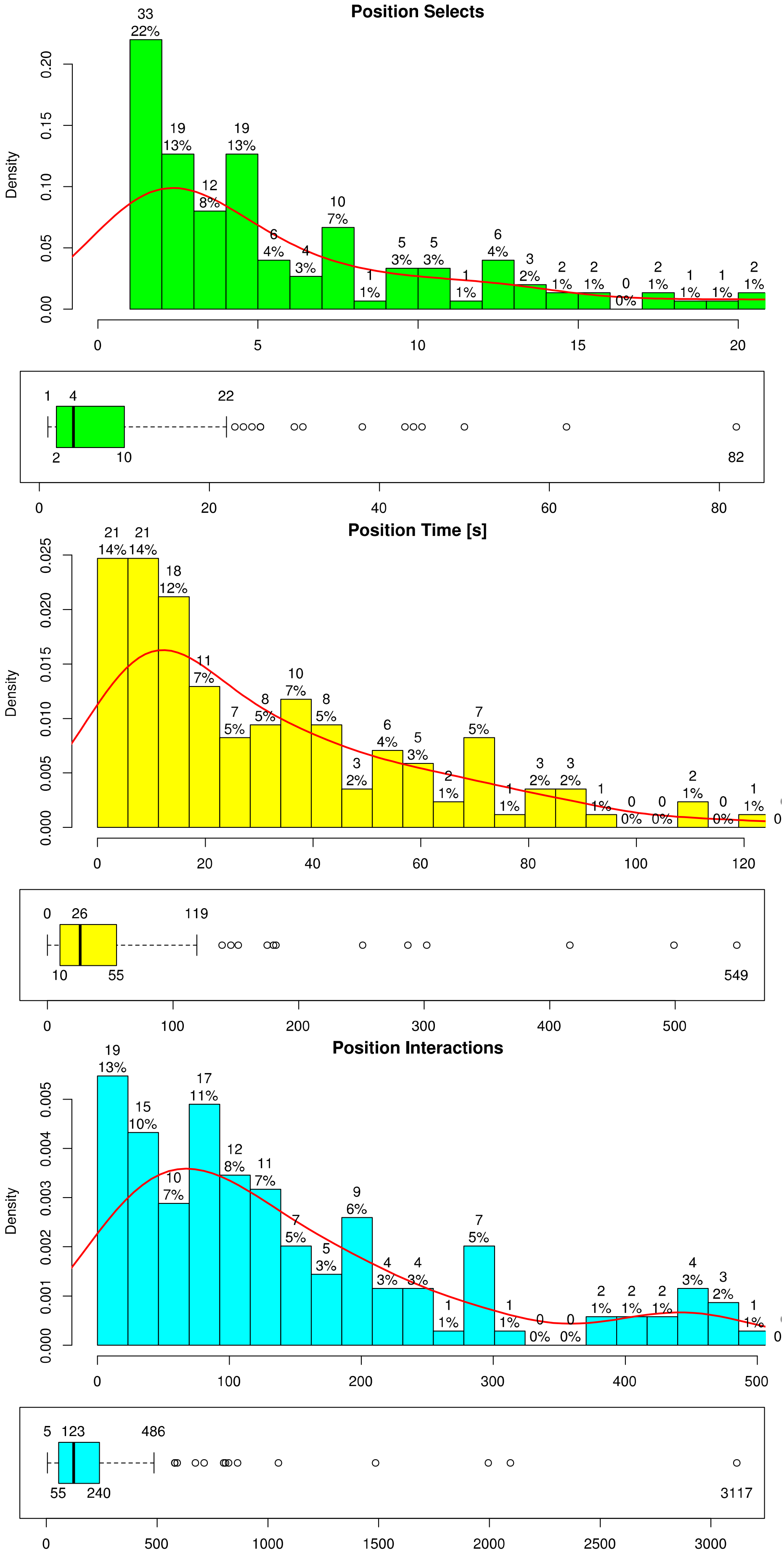


Figure 6.16: Selects, Time and Interactions to Select the requested Position

The middle 50 per cent of research subjects solved the task by performing 2 to 10 selects and the median by performing 4 selects. The most effective respondents solved the exercise by performing only 1 select and the participants less effective than any other, other than the outliers who might not participated seriously and performed up to 82 selects, with 22 selects. The greatest band of persons concerned within the detailed distribution, namely 22 per cent of them, solved the task by performing only 1 select.

The middle 50 per cent of research subjects solved the task within 10 to 55 seconds and the median in 26 seconds. The fastest respondents solved the exercise in less than 1 seconds and the participants slower than any other, other than the outliers who might not participated seriously and took up to 549 seconds, in 119 seconds. The greatest band of persons concerned within the detailed distribution, namely 28 per cent of them, solved the task within 0 and 12 seconds.

The middle 50 per cent of research subjects solved the task by performing 55 to 240 interactions and the median by performing 123 interactions. The most effective respondents solved the exercise by performing only 5 interactions and the participants less effective than any other, other than the outliers who might not participated seriously and performed up to 3117 interactions, with 486 interactions. The greatest band of persons concerned within the detailed distribution, namely 13 per cent of them, solved the task by performing between 0 and 23 interactions.

This confirms Hypothesis 2, which proposes that the user interface design concepts for presenting the position will be effective and enable blind and visually impaired persons to recognise the position of a document structure element or text node with a recall rate which will be much better than the number of selects required to recognise the position of a document structure element or text node by chance only, within a usable amount of time taken and with a usable effort of performed interactions.

The effectiveness is dependent on the input modality used by a specific person. The time taken depends to 4 per cent on the input modality, as shown by a weak bivariate correlation with $v = 0.20$ and a high level of statistical significance with $p = 0.014$. The number of interactions performed depends to 6 per cent on the input modality, as shown by a weak bivariate correlation with $v = 0.25$ and a high level of statistical significance with $p = 0.0019$. Figure 6.17 shows the central tendency, the median and dispersion as well as any outliers covering the entire time and number of interactions range in correlation to the different input modalities used:

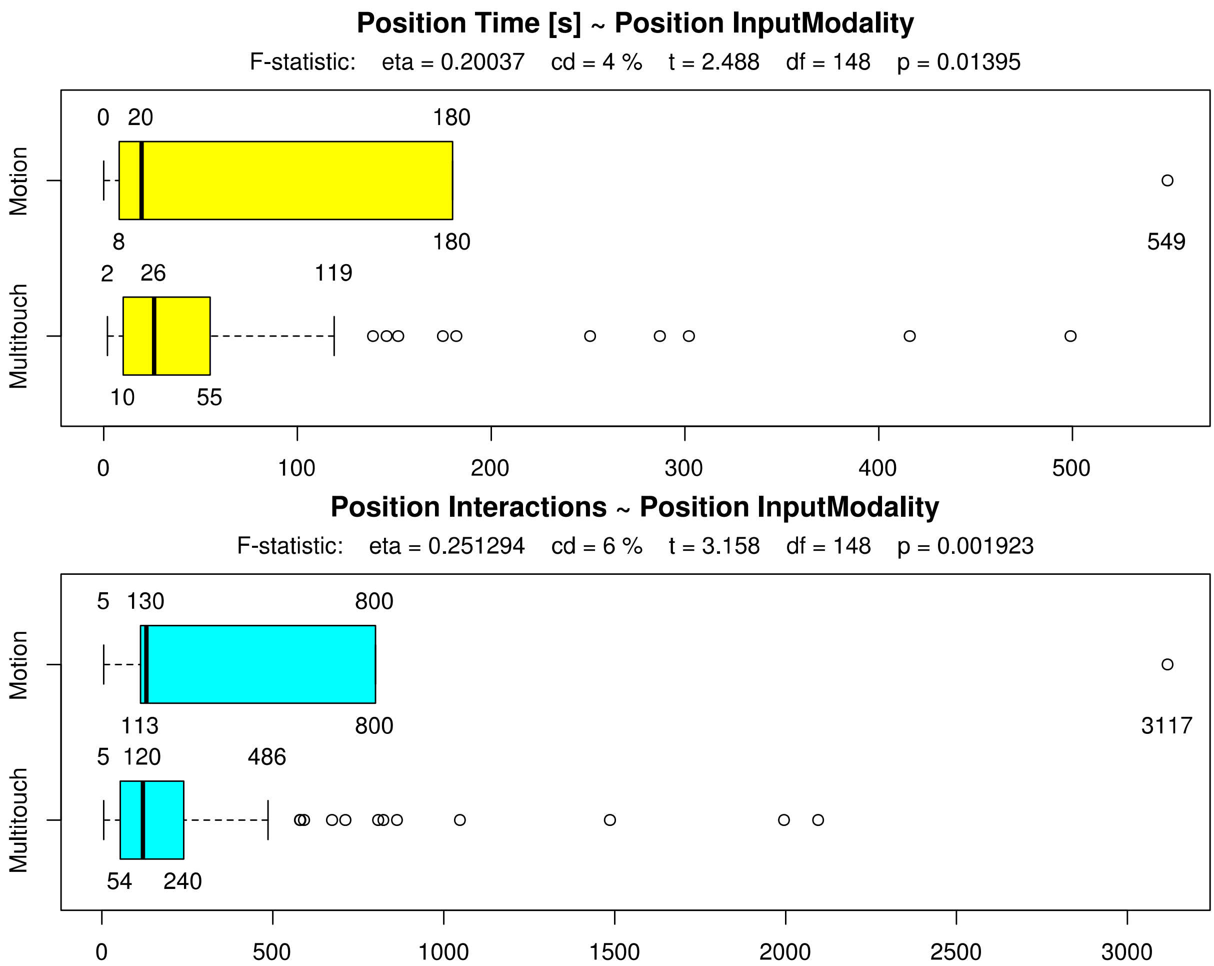


Figure 6.17: Position Time and Interactions in correlation to Input Modality

The middle 50 per cent of research subjects who used multitouch solved the task within 10 to 55 seconds and the median in 26 seconds. The fastest respondents solved the exercise in only 2 seconds and the participants slower than any other, other than the outliers, in 119 seconds. On the other hand, the middle 50 per cent of research subjects who used motion solved the task within 8 to 180 seconds and the median in 20 seconds. The fastest respondents solved the exercise in less than 1 second and the participants slower than any other, other than the outliers who took up to 549 seconds, in 180 seconds too.

The middle 50 per cent of research subjects who used multitouch solved the task by performing 54 to 240 interactions and the median by performing 120 interactions. The most effective respondents solved the exercise by performing only 5 interactions and the participants less effective than any other, other than the outliers, with 486 interactions. On the other hand, the middle 50 per cent of research subjects who used motion solved the task by performing 113 to 800 interactions and the median by performing 130 interactions. The most effective respondents solved the exercise by performing only 5 interactions and the participants less effective than any other, other than the outliers who performed up to 3117 interactions, with 800 interactions too.

Therefore Hypothesis 3, which proposes that the same effectiveness (number of selects performed, time taken and number of interactions performed) will be achieved across all different input modalities to find and select the document structure element at the requested position, can be rejected.

Hypothesis 4, which proposes that blind and visually impaired persons who are using screen reader will perform better than respondents who are not using screen reader, can be rejected because no statistically significant correlation between the use of screen reader and the number of selects performed, time taken and number of interactions performed to find and select the document structure element at the requested position could be found.

Hypothesis 5, which proposes that the same effectiveness will be achieved across all the different hardware devices employed, can be confirmed because no statistically significant correlation between the hardware device used and the number of selects performed, time taken and number of interactions performed to find and select the document structure element at the requested position could be found.

There is a multivariate relationship between the number of selects performed, the time taken and the number of interactions performed to find and select the document structure element at the requested position. The number of selects performed depends to 19 per cent on the time taken to solve the task, as shown by a moderate bivariate correlation with cor = 0.43 and a very high level of statistical significance with p = 2.93e-8 as well as to 23 percent on the number of interactions performed, as shown by another moderate bivariate correlation with cor = 0.48 and a very high statistical significance level of p = 4.91e-10. The time taken itself depends to 83 per cent on the number of interactions performed to solve the task, as shown by a very strong bivariate correlation with cor = 0.91 an a very high level of statistical significance with p = 0. Figure 6.18 shows the number of selects performed in correlation to the time taken and the number of interactions performed as well as the regression pane of this relationship:

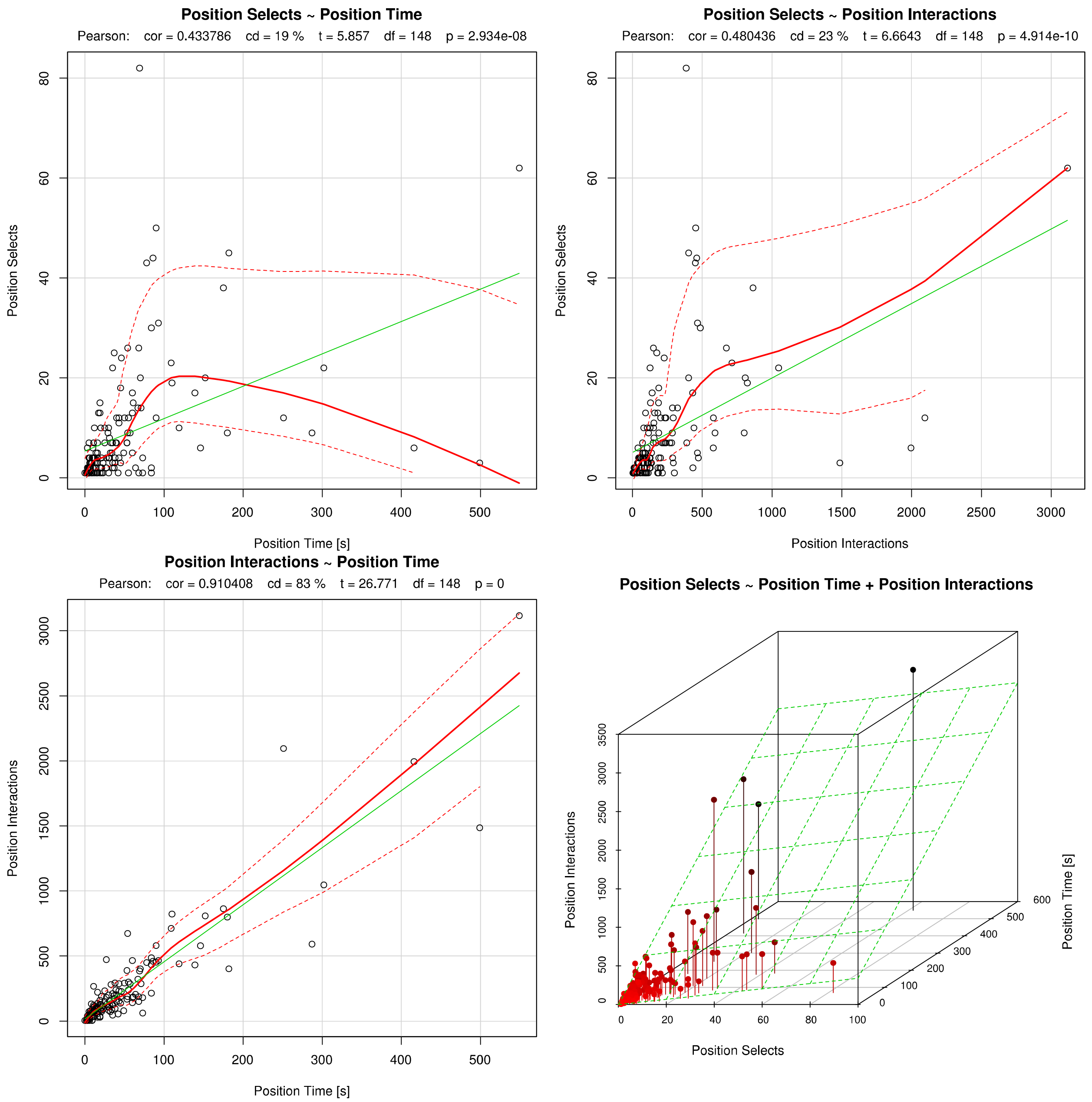


Figure 6.18: Position Selects in correlation to Time and Interactions

This correlation shows that the more selects had been performed, the more time was taken and the more interactions had been performed. These results confirm Hypothesis 6, which proposes that the more effective the user interface design concepts for presenting the position are, the less selects will be performed, the less time will be taken and the less interactions will be performed.

6.3.4 Length

In this exercise, the participants had to find and select the document structured element containing the longest text in order to examine if and how well the respondents are able to recognise the length of a document structure element or text node. The length is presented to the user by the length on the vertical y-axis of the multi-touch screen or multi-touch pad of the mobile and wearable device from the top to the bottom and the position on the horizontal orientation axis of the device anticlockwise within an angle of 180 degrees.

The research subjects used different input modalities to find and select the document structure element with the requested length. Figure 6.19 shows the different input modalities used by the persons concerned as well as the frequencies and percentages of respondents using each input modality:

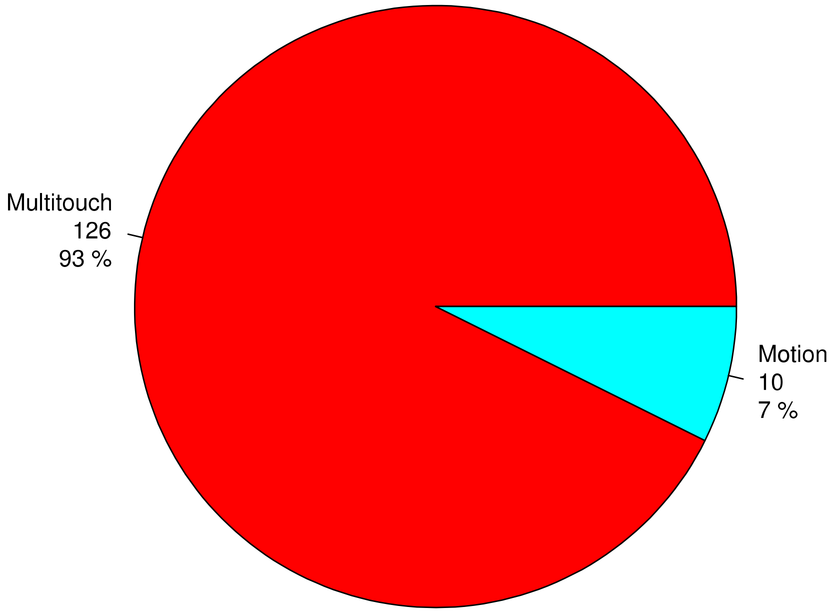


Figure 6.19: Input Modality used to Select the requested Length

Most of the research subjects, namely 93 per cent of them, used multitouch and only 7 per cent of the persons concerned used motion to find and select the document structure element with the requested length. Hypothesis 1, which proposes that blind and visually impaired persons who are using mobile devices like smart-phones and smart-tablets will choose to use the multitouch input modality to recognise the length of a document structure element or text node, can be confirmed.

The number of selects performed, time taken and number of interactions performed by the research subjects to find and select the document structure element with the requested length greatly differs between the cases. Figure 6.20 shows the central tendency, the median and dispersion as well as any outliers covering the entire number of selects, time and number of interactions range as well as frequencies and percentages of the detailed distribution of the interesting ranges only without the outliers:

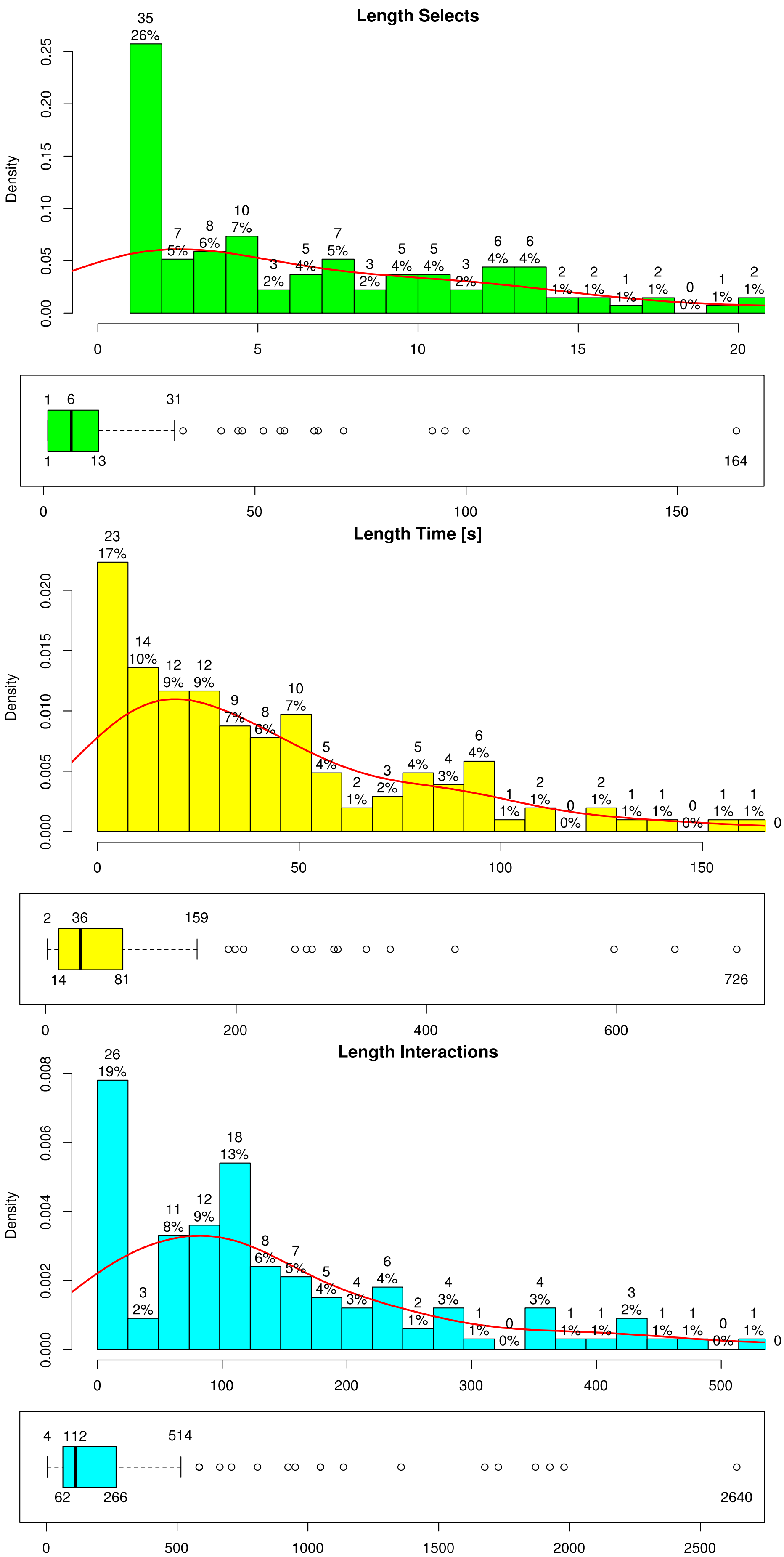


Figure 6.20: Selects, Time and Interactions to Select the requested Length

The middle 50 per cent of research subjects solved the task by performing 1 to 13 selects and the median by performing 6 selects. The most effective respondents solved the exercise by performing only 1 select and the participants less effective than any other, other than the outliers who might not participated seriously and performed up to 164 selects, with 31 selects. The greatest band of persons concerned within the detailed distribution, namely 26 per cent of them, solved the task by performing only 1 select.

The middle 50 per cent of research subjects solved the task within 14 to 81 seconds and the median in 36 seconds. The fastest respondents solved the exercise in 2 seconds and the participants slower than any other, other than the outliers who might not participated seriously and took up to 726 seconds, in 159 seconds. The greatest band of persons concerned within the detailed distribution, namely 17 per cent of them, solved the task within 0 and 8 seconds.

The middle 50 per cent of research subjects solved the task by performing 62 to 266 interactions and the median by performing 112 interactions. The most effective respondents solved the exercise by performing only 4 interactions and the participants less effective than any other, other than the outliers who might not participated seriously and performed up to 2640 interactions, with 514 interactions. The greatest band of persons concerned within the detailed distribution, namely 19 per cent of them, solved the task by performing between 0 and 25 interactions.

This confirms Hypothesis 2, which proposes that the user interface design concepts for presenting the length will be effective and enable blind and visually impaired persons to recognise the length of a document structure element or text node with a recall rate which will be much better than the number of selects required to recognise the length of a document structure element or text node by chance only, within a usable amount of time taken and with a usable effort of performed interactions.

The effectiveness is dependent on the input modality used by a specific person. The time taken depends to 3 per cent on the input modality, as shown by a weak bivariate correlation with $v = 0.18$ and a high level of statistical significance with $p = 0.036$. The number of interactions performed depends to 3 per cent on the input modality, as shown by a weak bivariate correlation with $v = 0.18$ and a high level of statistical significance with $p = 0.041$. Figure 6.21 shows the central tendency, the median and dispersion as well as any outliers covering the entire time and number of interactions range in correlation to the different input modalities used:

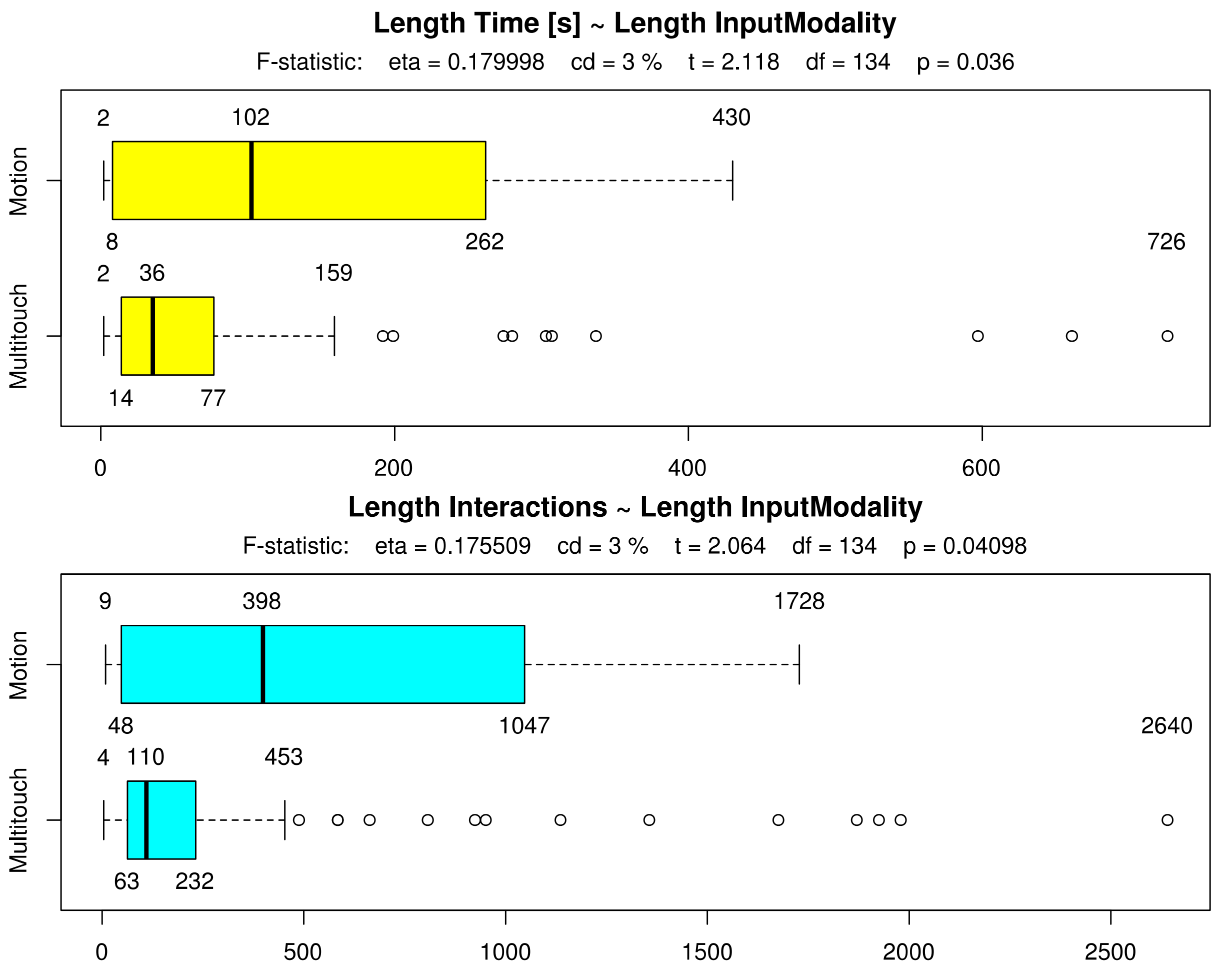


Figure 6.21: Length Time and Interactions in correlation to Input Modality

The middle 50 per cent of research subjects who used multitouch solved the task within 14 to 77 seconds and the median in 36 seconds. The fastest respondents solved the exercise in 2 seconds and the participants slower than any other, other than the outliers, in 159 seconds. On the other hand, the middle 50 per cent of research subjects who used motion solved the task within 8 to 262 seconds and the median in 102 seconds. The fastest respondents solved the exercise in 2 seconds and the participants slower than any other, other than the outliers who took up to 726 seconds, in 430 seconds.

The middle 50 per cent of research subjects who used multitouch solved the task by performing 63 to 232 interactions and the median by performing 110 interactions. The most effective respondents solved the exercise by performing only 4 interactions and the participants less effective than any other, other than the outliers who performed up to 2640 interactions, with 453 interactions. On the other hand, the middle 50 per cent of research subjects who used motion solved the task by performing 48 to 1047 interactions and the median by performing 398 interactions. The most effective respondents solved the exercise by performing only 9 interactions and the participants less effective than any other, other than the outliers, with 1728 interactions.

Therefore Hypothesis 3, which proposes that the same effectiveness (number of selects performed, time taken and number of interactions performed) will be achieved across all different input modalities to find and select the document structure element with the requested length, can be rejected.

Hypothesis 4, which proposes that blind and visually impaired persons who are using screen reader will perform better than respondents who are not using screen reader, can be rejected because no statistically significant correlation between the use of screen reader and the number of selects performed, time taken and number of interactions performed to find and select the document structure element with the requested length could be found.

Hypothesis 5, which proposes that the same effectiveness will be achieved across all the different hardware devices employed, can be confirmed because no statistically significant correlation between the hardware device used and the number of selects performed, time taken and number of interactions performed to find and select the document structure element with the requested length could be found.

There is a multivariate relationship between the number of selects performed, the time taken and the number of interactions performed to find and select the document structure element with the requested length. The number of selects performed depends to 31 per cent on the time taken to solve the task, as shown by a moderate bivariate correlation with cor = 0.55 and a very high level of statistical significance with $p = 2.41e-12$ as well as to 44 percent on the number of interactions performed, as shown by strong bivariate correlation with cor = 0.66 and a very high statistical significance level of $p = 0$. The time taken itself depends to 88 per cent on the number of interactions performed to solve the task, as shown by a very strong bivariate correlation with cor = 0.94 an a very high level of statistical significance with $p = 0$. Figure 6.22 shows the number of selects performed in correlation to the time taken and the number of interactions performed as well as the regression pane of this relationship:

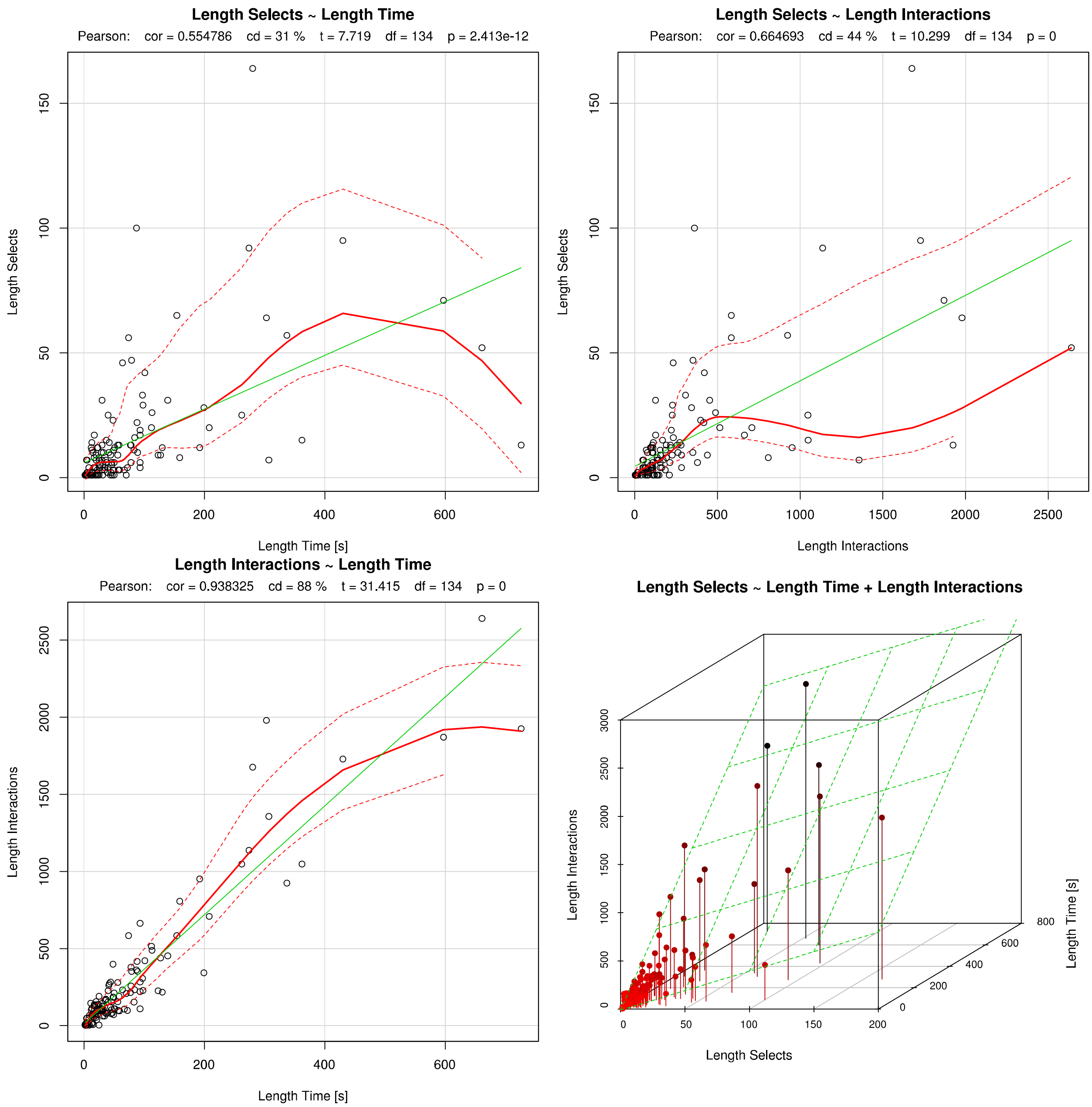


Figure 6.22: Length Selects in correlation to Time and Interactions

This correlation shows that the more selects had been performed, the more time was taken and the more interactions had been performed. These results confirm Hypothesis 6, which proposes that the more effective the user interface design concepts for presenting the length are, the less selects will be performed, the less time will be taken and the less interactions will be performed.

6.3.5 Relationship

In this exercise, the participants had to find and select the paragraph which is inside 3 sections in order to examine if and how well the respondents are able to recognise the relationship of all nested document structure element or text nodes at a position. The relationship is presented to the user by rendering their single-pitch inherited elementary earcons simultaneously as parallel compound earcons to reduce the length of time this compound audio message takes, rendering their single-motive inherited elementary tactons sequentially as serial compound tactons and by rendering their inherited elementary speech utterances sequentially as serial compound utterances.

The research subjects used different input modalities to find and select the document structure element with the requested relationship. Figure 6.23 shows the different input modalities used by the persons concerned as well as the frequencies and percentages of respondents using each input modality:

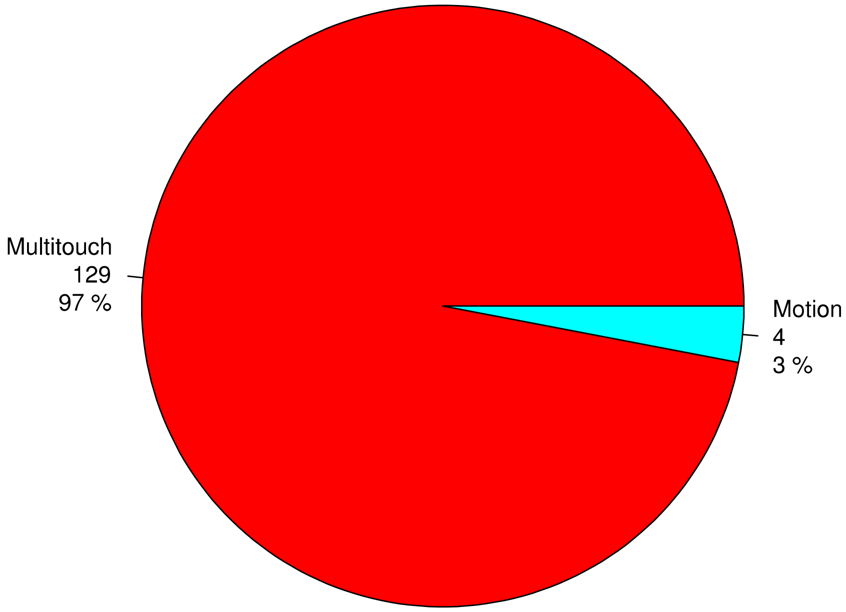


Figure 6.23: Input Modality used to Select the requested Relationship

Most of the research subjects, namely 97 per cent of them, used multitouch and only 3 per cent of the persons concerned used motion to find and select the document structure element with the requested relationship. Hypothesis 1, which proposes that blind and visually impaired persons who are using mobile devices like smart-phones and smart-tablets will choose to use the multitouch input modality to recognise the relationship of all nested document structure element or text nodes at a position, can be confirmed.

The number of selects performed, time taken and number of interactions performed by the research subjects to find and select the document structure element with the requested relationship greatly differs between the cases. Figure 6.24 shows the central tendency, the median and dispersion as well as any outliers covering the entire number of selects, time and number of interactions range as well as frequencies and percentages of the detailed distribution of the interesting ranges only without the outliers:

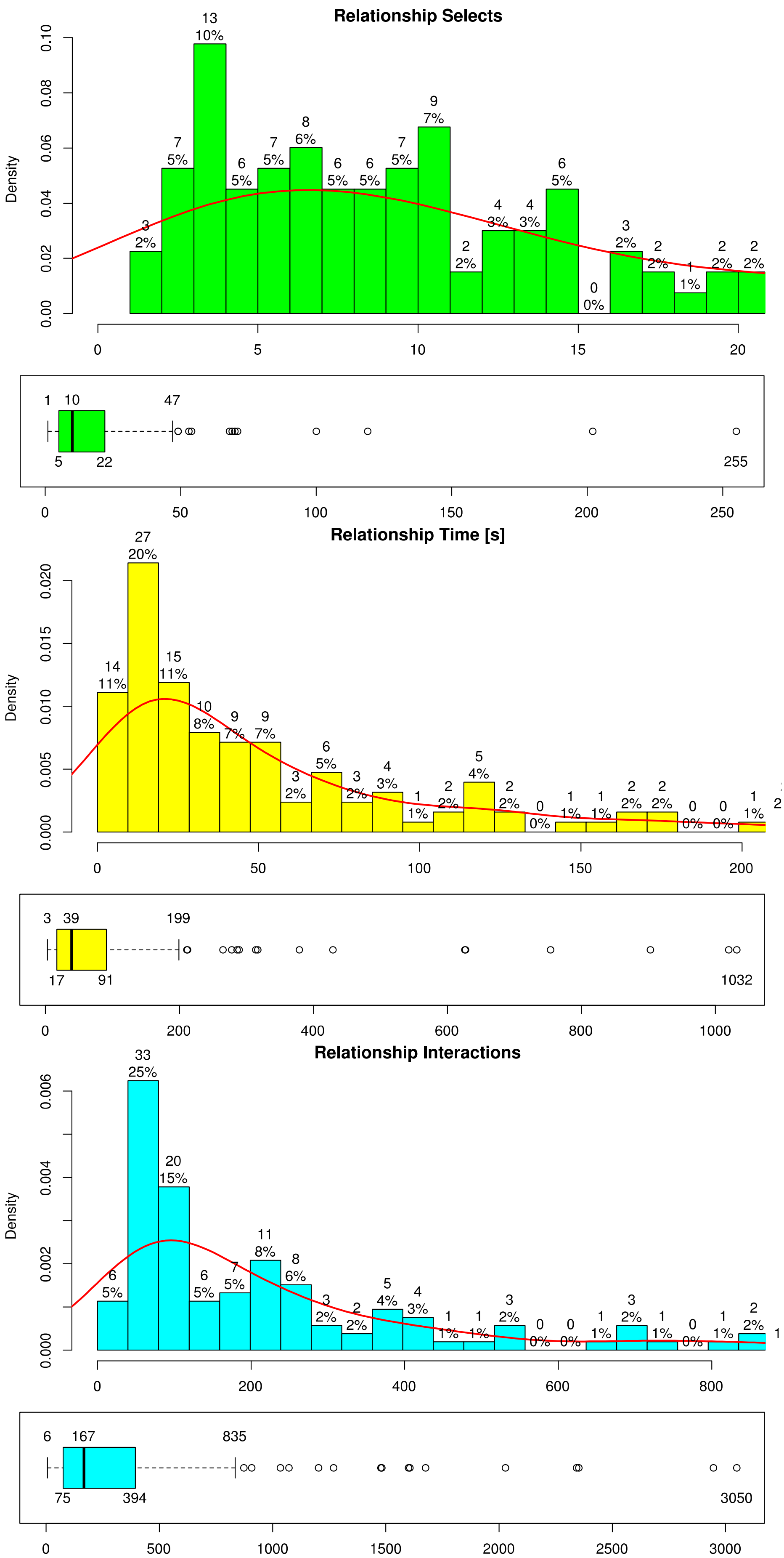


Figure 6.24: Selects, Time and Interactions to Select the requested Relationship

The middle 50 per cent of research subjects solved the task by performing 5 to 22 selects and the median by performing 10 selects. The most effective respondents solved the exercise by performing only 1 select and the participants less effective than any other, other than the outliers who might not participated seriously and performed up to 255 selects, with 47 selects. The greatest band of persons concerned within the detailed distribution, namely 10 per cent of them, solved the task by performing 3 selects.

The middle 50 per cent of research subjects solved the task within 17 to 91 seconds and the median in 39 seconds. The fastest respondents solved the exercise in 3 seconds and the participants slower than any other, other than the outliers who might not participated seriously and took up to 1032 seconds, in 199 seconds. The greatest band of persons concerned within the detailed distribution, namely 20 per cent of them, solved the task within 10 and 20 seconds.

The middle 50 per cent of research subjects solved the task by performing 75 to 394 interactions and the median by performing 167 interactions. The most effective respondents solved the exercise by performing only 6 interactions and the participants less effective than any other, other than the outliers who might not participated seriously and performed up to 3050 interactions, with 835 interactions. The greatest band of persons concerned within the detailed distribution, namely 25 per cent of them, solved the task by performing between 40 and 80 interactions.

These results fit in well with the empirically derived guidelines for the presentation of concurrent earcons of McGookin & Brewster (2004) where the identification of earcons fell from 70 per cent, consistent with similar work by Brewster (1994) on single earcon identification, when only one earcon was presented, to 30 per cent when four earcons were concurrently presented (McGookin & Brewster, 2003). Identification of both timbre and rhythm encoded data attributes is much higher, dropping from 95 per cent to 65 per cent as the number of concurrently presented earcons is increased. Identification of the register encoded earcon attribute, whilst being much lower than timbre and rhythm, does fall at a much shallower gradient, due in part to the incorporation of inharmonic musical intervals between the registers used (McGookin & Brewster, 2004).

In addition, some research subjects complained that they feel annoyed by the continuous sounds. In this cases, the continuous sounds failed to fade into the background of consciousness after a short period of time as proposed by auditory habituation (Buxton, Gaver & Bly, 1991), which is one important ability of the human auditory system, and come to the foreground of attention If the sound was to change or stop because of the sensitivity of the auditory system to changes in stimulus (Buxton, 1989).

The serial compound tactons might not be able to keep up with the pace of all interactions that are going on in the user interface at the same time because they can take a long time to play since each motive lasts a particular length of time depending on its notes and the tempo and these are then combined to produce longer compound tactons. They could be played more rapidly (at a faster tempo) to overcome this problem but then errors in recognition may occur.

Therefore Hypothesis 2, which proposes that the user interface design concepts for presenting the relationship will be effective and enable blind and visually impaired persons to recognise the relationship of all nested document structure element or text nodes at a position with a recall rate which will be much better than the number of selects required to recognise the relationship of all nested document structure element or text nodes at a position by chance only, within a usable amount of time taken and with a usable effort of performed interactions, can be partially rejected.

Hypothesis 3, which proposes that the same effectiveness (number of selects performed, time taken and number of interactions performed) will be achieved across all different input modalities, can be confirmed because no statistically significant correlation between the input modality used and the number of selects performed, time taken and number of interactions performed to find and select the document structure element with the requested relationship could be found.

The effectiveness is dependent on the use of screen reader by a specific person. The number of selects performed depends to 3 per cent on the use of screen reader, as shown by a weak bivariate correlation with v = 0.16 and a high level of statistical significance with p = 0.062. Figure 6.25 shows the central tendency, the median and dispersion as well as any outliers covering the entire number of selects range in correlation to the use of screen reader:

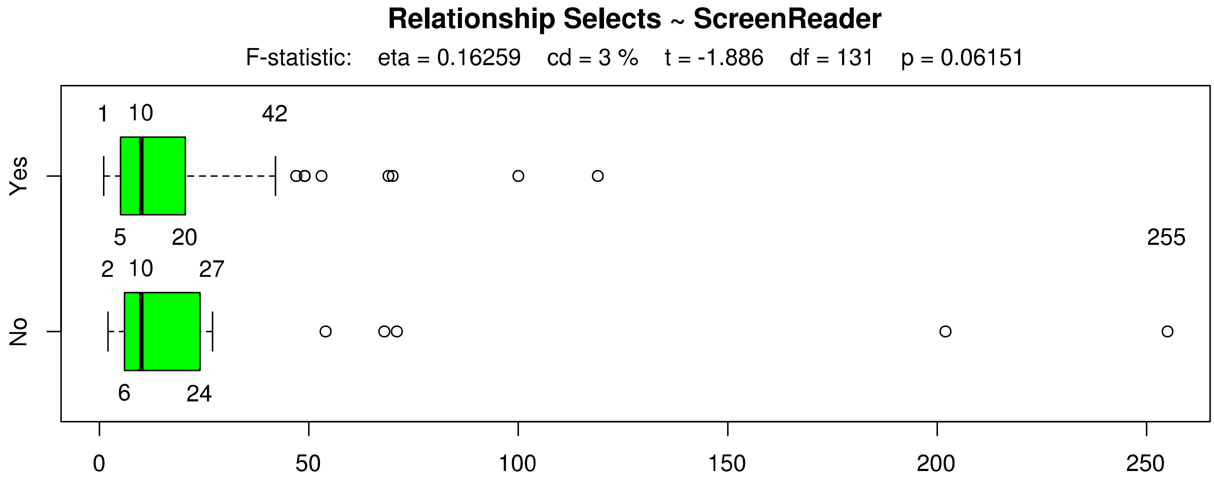


Figure 6.25: Relationship Selects in correlation to use of Screen Reader

The middle 50 per cent of research subjects who are using screen reader solved the task by performing 5 to 20 selects and the median by performing 10 selects. The most effective respondents solved the exercise by performing only 1 select and the participants less effective than any other, other than the outliers, with 42 selects. On the other hand, the middle 50 per cent of research subjects who are not using screen reader solved the task by performing 6 to 24 selects and the median by performing 10 selects. The most effective respondents solved the exercise by performing only 2 selects and the participants less effective than any other, other than the outliers who performed up to 255 selects, with 27 selects.

Therefore Hypothesis 4, which proposes that blind and visually impaired persons who are using screen reader will perform better than respondents who are not using screen reader to find and select the document structure element with the requested relationship, can be confirmed.

Hypothesis 5, which proposes that the same effectiveness will be achieved across all the different hardware devices employed, can be confirmed because no statistically significant correlation between the hardware device used and the number of selects performed, time taken and number of interactions performed to find and select the document structure element with the requested relationship could be found.

There is a multivariate relationship between the number of selects performed, the time taken and the number of interactions performed to find and select the document structure element with the requested relationship. The number of selects performed depends to 20 per cent on the time taken to solve the task, as shown by a moderate bivariate correlation with cor = 0.45 and a very high level of statistical significance with p = 6.65e-8 as well as to 30 percent on the number of interactions performed, as shown by another moderate bivariate correlation with cor = 0.55 and a very high statistical significance level of p = 8.71e-12. The time taken itself depends to 86 per cent on the number of interactions performed to solve the task, as shown by a very strong bivariate correlation with cor = 0.93 an a very high level of statistical significance with p = 0. Figure 6.26 shows the number of selects performed in correlation to the time taken and the number of interactions performed as well as the regression pane of this relationship:

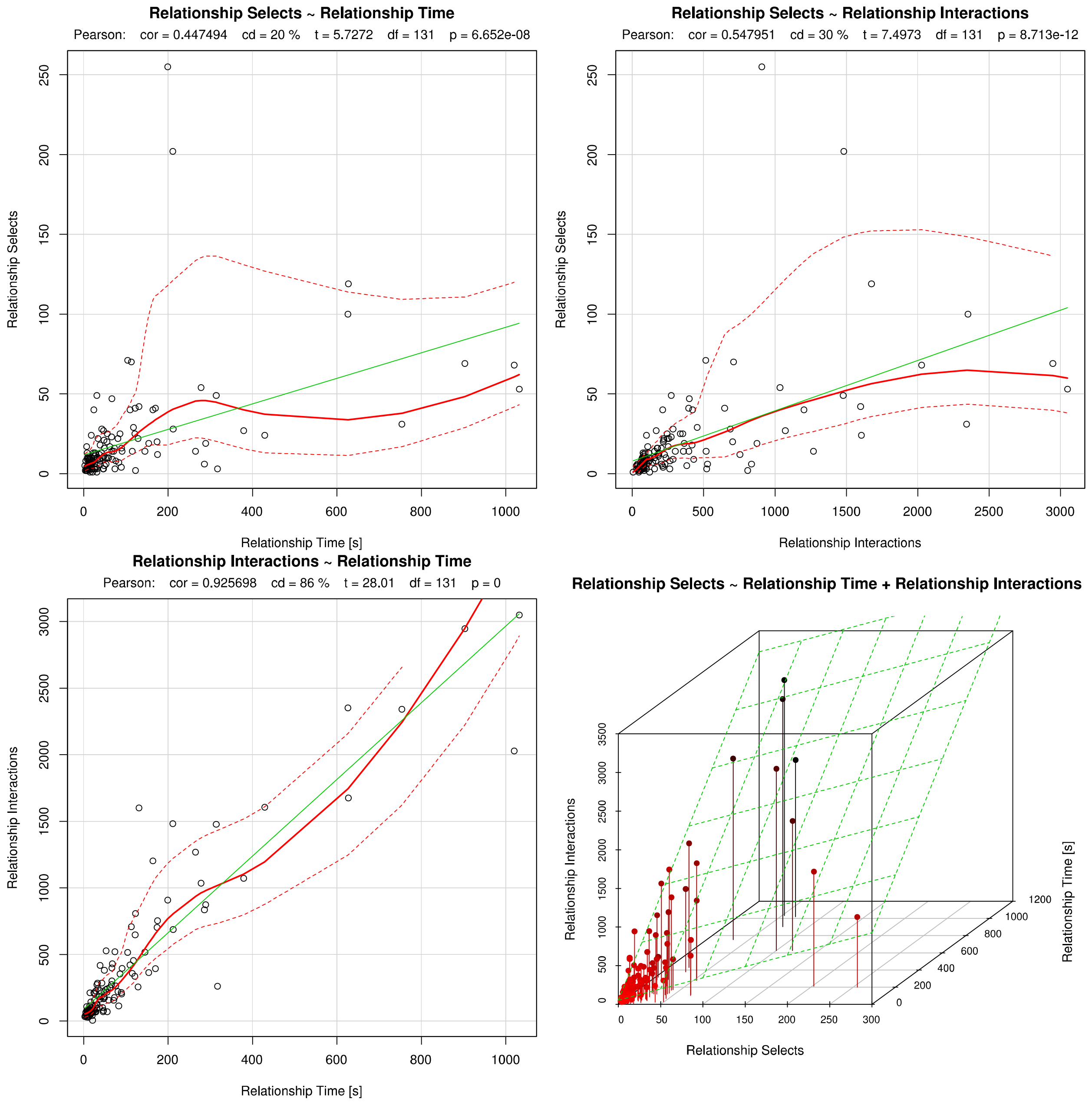


Figure 6.26: Relationship Selects in correlation to Time and Interactions

This correlation shows that the more selects had been performed, the more time was taken and the more interactions had been performed. These results confirm Hypothesis 6, which proposes that the more effective the user interface design concepts for presenting the relationship are, the less selects will be performed, the less time will be taken and the less interactions will be performed.

6.3.6 Text

In this exercise, the participants had to find and select the element which contains the text "Strawberry" in order to examine if and how well the respondents are able to recognise the text contained in a document structure element or text node. The text is presented to the user by the text of the inherited elementary speech utterance.

The research subjects used different input modalities to find and select the document structure element with the requested text. Figure 6.27 shows the different input modalities used by the persons concerned as well as the frequencies and percentages of respondents using each input modality:

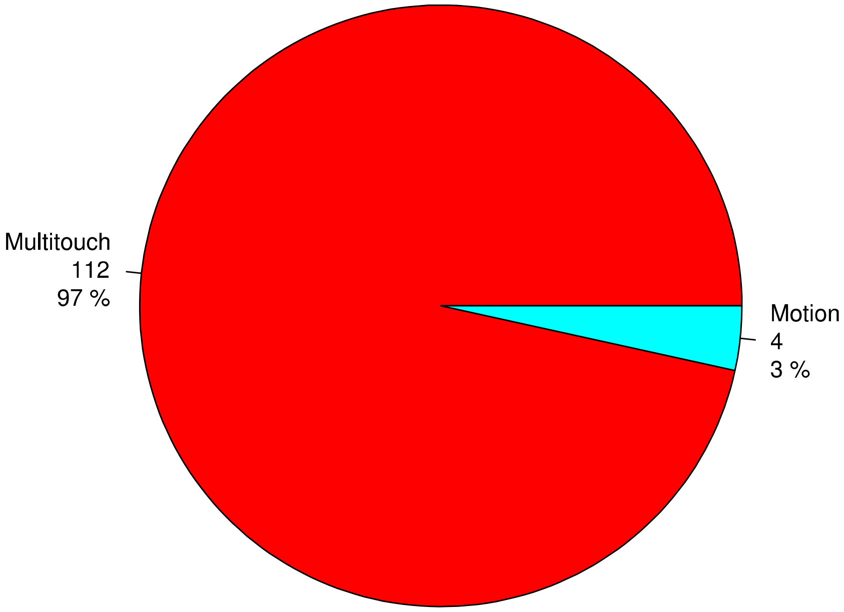


Figure 6.27: Input Modality used to Select the requested Text

Most of the research subjects, namely 97 per cent of them, used multitouch and only 3 per cent of the persons concerned used motion to find and select the document structure element with the requested text. Hypothesis 1, which proposes that blind and visually impaired persons who are using mobile devices like smart-phones and smart-tablets will choose to use the multitouch input modality to recognise the text contained in a document structure element or text node, can be confirmed.

The number of selects performed, time taken and number of interactions performed by the research subjects to find and select the document structure element with the requested text greatly differs between the cases. Figure 6.28 shows the central tendency, the median and dispersion as well as any outliers covering the entire number of selects, time and number of interactions range as well as frequencies and percentages of the detailed distribution of the interesting ranges only without the outliers:

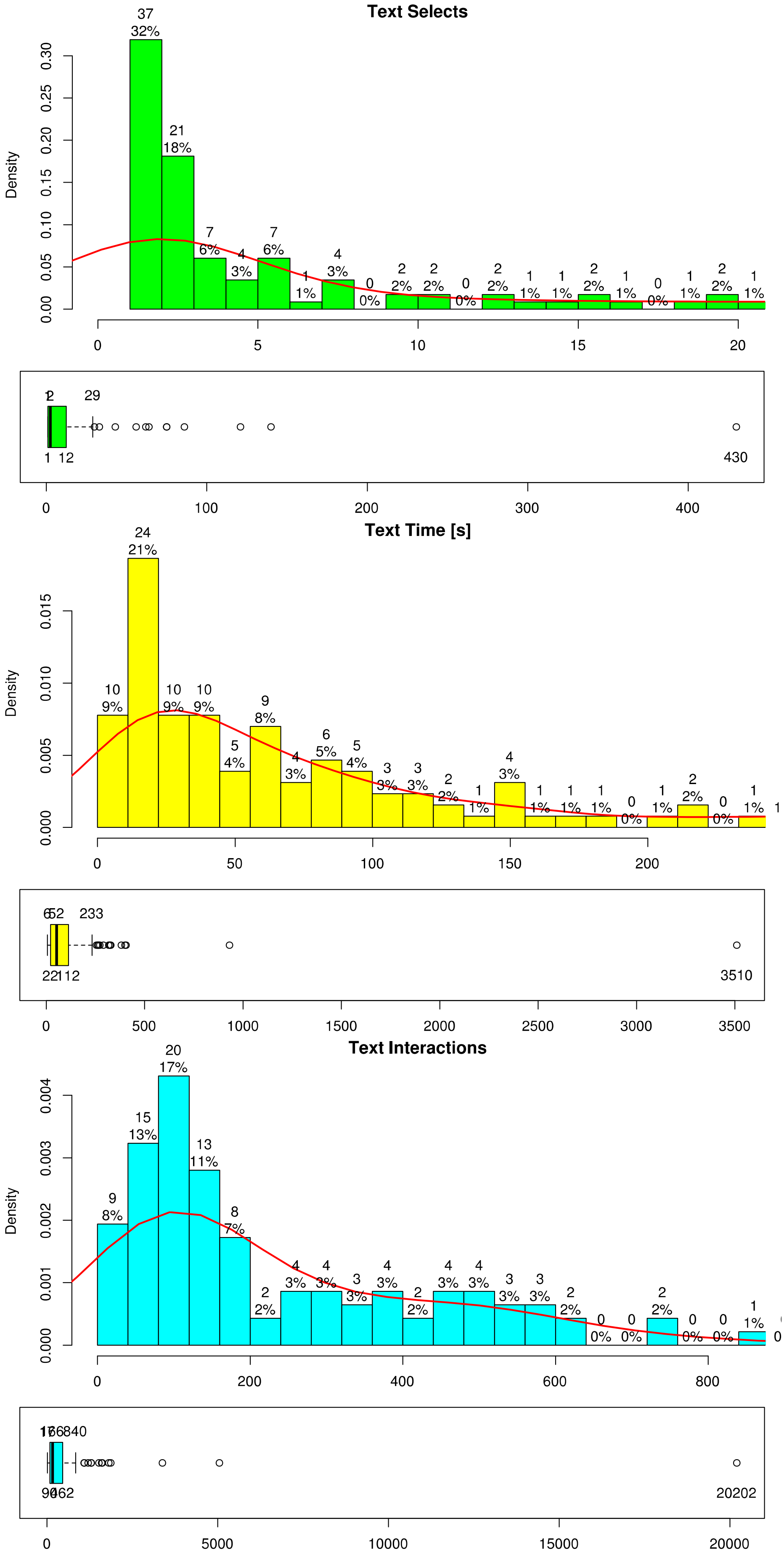


Figure 6.28: Selects, Time and Interactions to Select the requested Text

The middle 50 per cent of research subjects solved the task by performing 1 to 12 selects and the median by performing 2 selects. The most effective respondents solved the exercise by performing only 1 select and the participants less effective than any other, other than the outliers who might not participated seriously and performed up to 430 selects, with 29 selects. The greatest band of persons concerned within the detailed distribution, namely 32 per cent of them, solved the task by performing only 1 select.

The middle 50 per cent of research subjects solved the task within 22 to 112 seconds and the median in 52 seconds. The fastest respondents solved the exercise in 6 seconds and the participants slower than any other, other than the outliers who might not participated seriously and took up to 3510 seconds, in 233 seconds. The greatest band of persons concerned within the detailed distribution, namely 21 per cent of them, solved the task within 11 and 22 seconds.

The middle 50 per cent of research subjects solved the task by performing 75 to 394 interactions and the median by performing 167 interactions. The most effective respondents solved the exercise by performing only 6 interactions and the participants less effective than any other, other than the outliers who might not participated seriously and performed up to 3050 interactions, with 835 interactions. The greatest band of persons concerned within the detailed distribution, namely 25 per cent of them, solved the task by performing between 40 and 80 interactions.

This confirms Hypothesis 2, which proposes that the user interface design concepts for presenting the text will be effective and enable blind and visually impaired persons to recognise the text contained in a document structure element or text node with a recall rate which will be much better than the number of selects required to recognise the text of a document structure element or text node by chance only, within a usable amount of time taken and with a usable effort of performed interactions.

The effectiveness depends on the input modality used by a specific person. The number of selects performed depends to 22 per cent on the input modality, as shown by a moderate bivariate correlation with $v = 0.47$ and very a high level of statistical significance with $p = 9.89e\text{-}8$. The time taken depends to 23 per cent on the input modality, as shown by a moderate bivariate correlation with $v = 0.47$ and a very high level of statistical significance with $p = 7.44e\text{-}8$. The number of interactions performed depends to 29 per cent on the input modality, as shown by a moderate bivariate correlation with $v = 0.54$ and very a high level of statistical significance with $p = 5.38e\text{-}10$. Figure 6.29 shows the central tendency, the median and dispersion as well as any outliers covering the entire number of selects, time and number of interactions range in correlation to the different input modalities used:

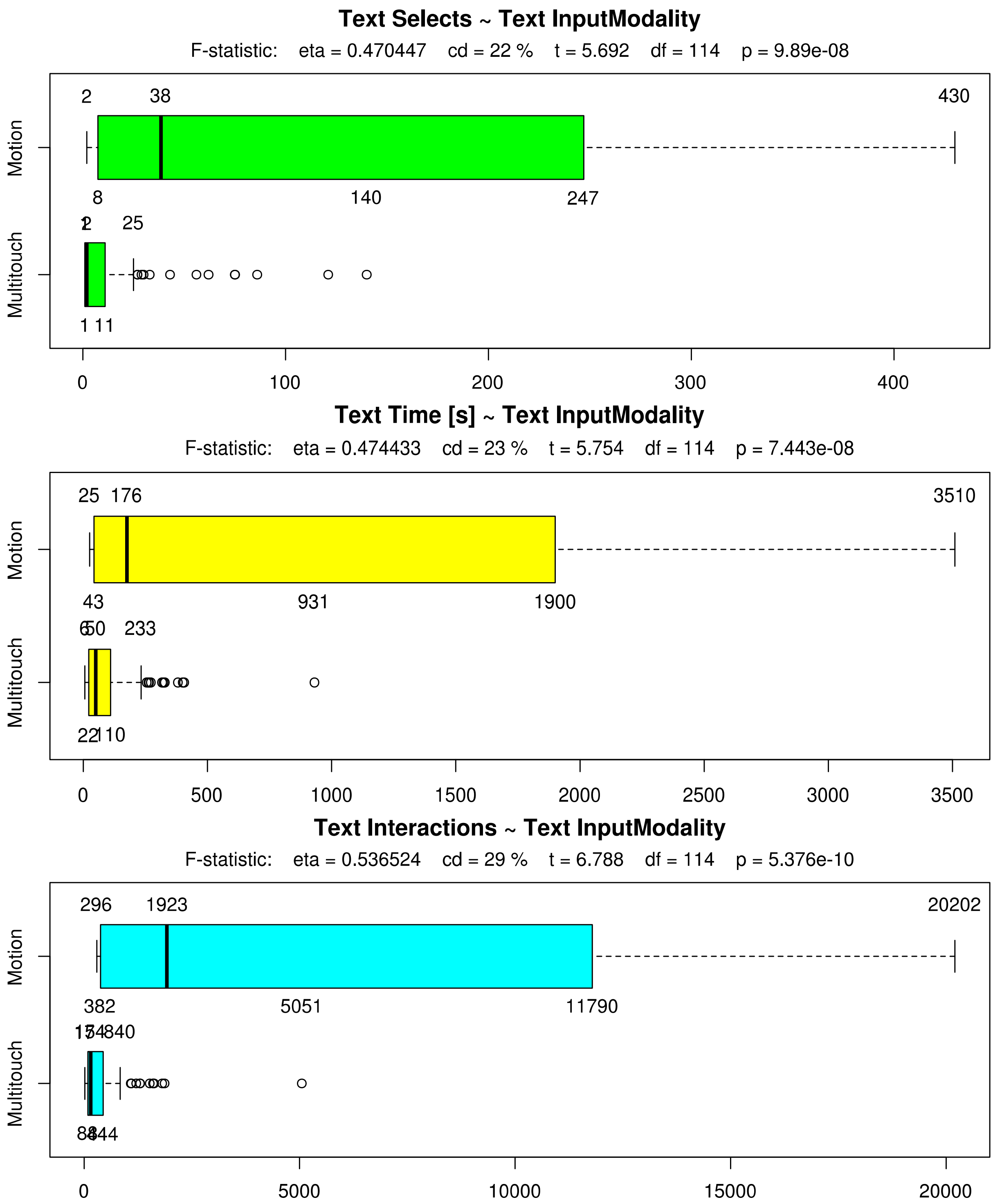


Figure 6.29: Text Selects, Time and Interaction in correlation to Input Modality

The middle 50 per cent of research subjects who solved the task by using a smart tablet performed 4 to 36 selects and the median performed 12 selects. The most effective respondents solved the exercise by performing only 2 selects and the participants less effective than any other, other than the outliers who performed up to 166 selects, with 67 selects. On the other hand, the middle 50 per cent of research subjects who solved the task by using a smart phone performed 2 to 16 selects and the median performed 5 selects. The most effective respondents solved the exercise by performing only 1 select and the participants less effective than any other, other than the outliers, with 30 selects.

The middle 50 per cent of research subjects who used multitouch solved the task within 22 to 110 seconds and the median in 50 seconds. The fastest respondents solved the exercise in 6 seconds and the participants slower than any other, other than the outliers, in 233 seconds. On the other hand, the middle 50 per cent of research subjects who used motion solved the task within 43 to 1900 seconds and the median in 176 seconds. The fastest respondents solved the exercise in 25 seconds and the participants slower than any other, other than the outliers, in 3510 seconds.

The middle 50 per cent of research subjects who used multitouch solved the task by performing 88 to 444 interactions and the median by performing 154 interactions. The most effective respondents solved the exercise by performing only 17 interactions and the participants less effective than any other, other than the outliers, with 840 interactions. On the other hand, the middle 50 per cent of research subjects who used motion solved the task by performing 382 to 11790 interactions and the median by performing 1923 interactions. The most effective respondents solved the exercise by performing only 296 interactions and the participants less effective than any other, other than the outliers, with 20202 interactions.

This correlation arises because there is no minimal unit size with the multitouch input modality and therefore document structure elements containing only a small portion of text may appear too small on the clock to be focused and selected reliable. To solve this problem, a minimal unit size which can be reliable focused should be introduced. Further research in the domain of motion gestures and psychophysics is required to find out how big this minimal unit size should be. Therefore Hypothesis 3, which proposes that the same effectiveness (number of selects performed, time taken and number of interactions performed) will be achieved across all different input modalities, can be rejected.

Therefore Hypothesis 3, which proposes that the same effectiveness (number of selects performed, time taken and number of interactions performed) will be achieved across all different input modalities to find and select the document structure element with the requested text, can be rejected.

Hypothesis 4, which proposes that blind and visually impaired persons who are using screen reader will perform better than respondents who are not using screen reader, can be rejected because no statistically significant correlation between the use of screen reader and the number of selects performed, time taken and number of interactions performed to find and select the document structure element with the requested text could be found.

Hypothesis 5, which proposes that the same effectiveness will be achieved across all the different hardware devices employed, can be confirmed because no statistically significant correlation between the hardware device used and the number of selects performed, time taken and number of interactions performed to find and select the document structure element with the requested text could be found.

There is a multivariate relationship between the number of selects performed, the time taken and the number of interactions performed to find and select the document structure element with the requested text. The number of selects performed depends to 88 per cent on the time taken to solve the task, as shown by a very strong bivariate correlation with $cor = 0.94$ and a very high level of statistical significance with $p = 0$ as well as to 88 per cent on the number of interactions performed, as shown by another very strong bivariate correlation with $cor = 0.94$ and a very high statistical significance level of $p = 0$. The time taken itself depends to 98 per cent on the number of interactions performed to solve the task, as shown by a very strong bivariate correlation with $cor = 0.99$ an a very high level of statistical significance with $p = 0$. Figure 6.30 shows the number of selects performed in correlation to the time taken and the number of interactions performed as well as the regression pane of this relationship:

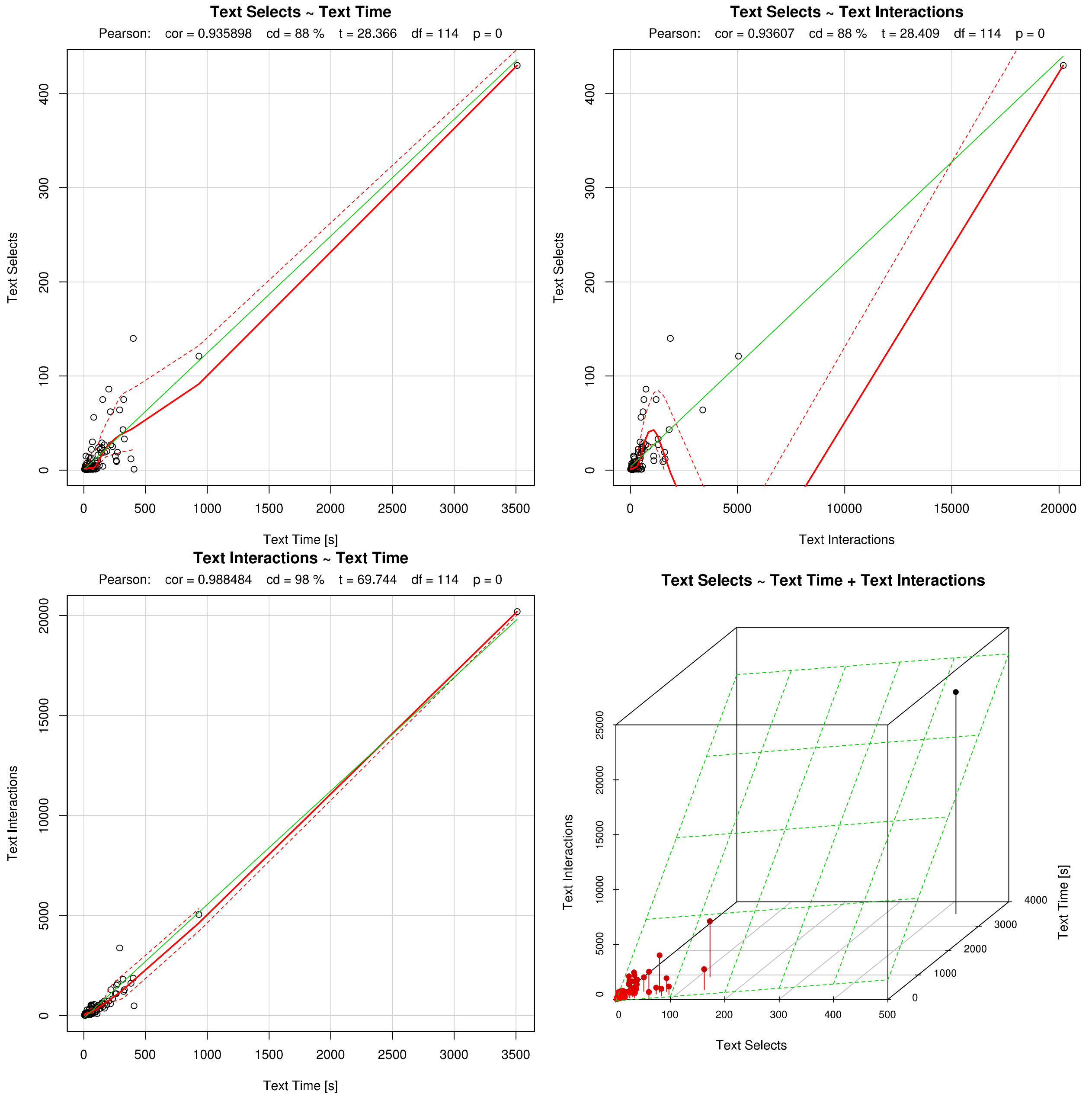


Figure 6.30: Text Selects in correlation to Time and Interactions

This correlation shows that the more selects had been performed, the more time was taken and the more interactions had been performed. These results confirm Hypothesis 6, which proposes that the more effective the user interface design concepts for presenting the text are, the less selects will be performed, the less time will be taken and the less interactions will be performed.

6.4 Navigation

The next 7 exercises focussing on the user interface design concepts and their hypotheses about the behaviour of blind and visually impaired persons among structured documents in relation to the actions for navigation provided by the different input modalities which are the Multitouch Input Modality as described in Chapter 5.2.5 and the Motion Input Modality as described in Chapter 5.2.6.

6.4.1 Set Structure Cursor

In this exercise, the participants had to set the structure cursor to an arbitrary position within the document in order to examine if and how well the respondents are able to setting the structure cursor to a position within the document. The structure cursor can be set by putting down one pointer at a specific position or by rotating the device horizontally anticlockwise 10 degrees.

The research subjects used different input modalities for setting the structure cursor. Figure 6.31 shows the different input modalities used by the persons concerned as well as the frequencies and percentages of respondents using each input modality:

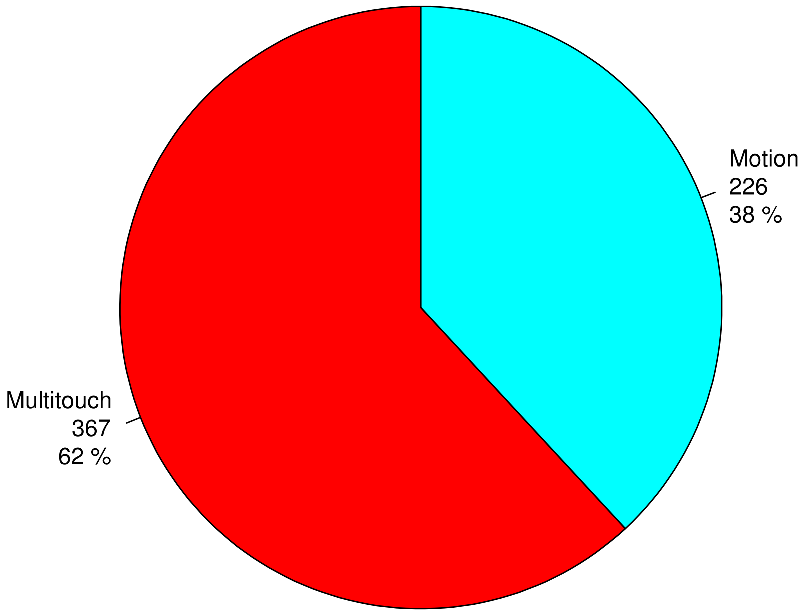


Figure 6.31: Input Modality used to Set the Structure Cursor

Most of the research subjects, namely 62 per cent of them, used multitouch and only 38 per cent of the persons concerned used motion for setting the structure cursor. Hypothesis 1, which proposes that blind and visually impaired persons who are using mobile devices like smart-phones and smart-tablets will choose to use the multitouch input modality for setting the structure cursor, can be confirmed.

The time taken by the research subjects for setting the structure cursor greatly differs between the cases. Figure 6.32 shows the central tendency, the median and dispersion as well as any outliers covering the entire time range as well as frequencies and percentages of the detailed distribution of the interesting time range only without the outliers:

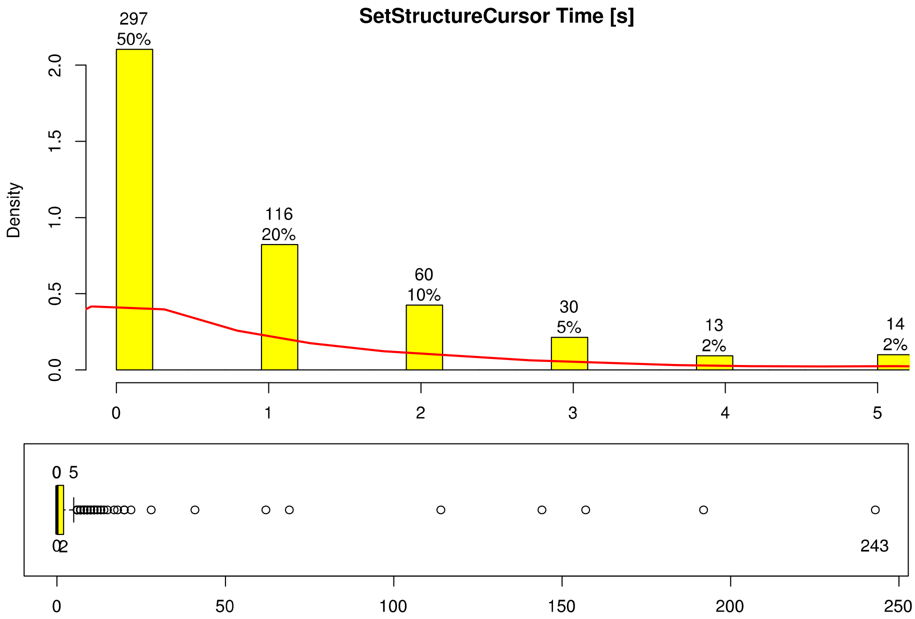


Figure 6.32: Time taken to Set the Structure Cursor

The middle 50 per cent of research subjects solved the task within 0 to 2 seconds and the median in less than 1 second too. The fastest respondents solved the exercise in less than 1 second and the participants slower than any other, other than the outliers who might not participated seriously and took up to 243 seconds, in 5 seconds. The greatest band of persons concerned within the detailed distribution, namely 50 per cent of them, solved the task in less than 1 second.

This confirms Hypothesis 2, which proposes that the user interface design concepts for setting the structure cursor will be effective and enable blind and visually impaired persons to set the structure cursor to an arbitrary position within the document within a usable amount of time taken.

Hypothesis 3, which proposes that the same effectiveness (time taken) will be achieved across all different input modalities, can be confirmed because no statistically significant correlation between the input modality used and the time taken to setting the structure cursor could be found.

Hypothesis 4, which proposes that blind and visually impaired persons who are using screen reader will perform better than respondents who are not using screen reader, can be rejected because no statistically significant correlation between the use of screen reader and the time taken to setting the structure cursor could be found.

Hypothesis 5, which proposes that the same effectiveness will be achieved across all the different hardware devices employed, can be confirmed because no statistically significant correlation between the hardware device used and the time taken to setting the structure cursor could be found.

6.4.2 Move Structure Cursor

In this exercise, the participants had to move the structure cursor from the beginning to the end of the example document in order to examine if and how well the respondents are able to moving the structure cursor forward and backward in order to obtain a fast overview over the structure of the entire document. The structure cursor can be moved by moving the one pointer up and down at an arbitrary speed or by rotating the device horizontally clockwise or anticlockwise within an angle of 180 degrees.

The research subjects used different input modalities for moving the structure cursor. Figure 6.33 shows the different input modalities used by the persons concerned as well as the frequencies and percentages of respondents using each input modality:

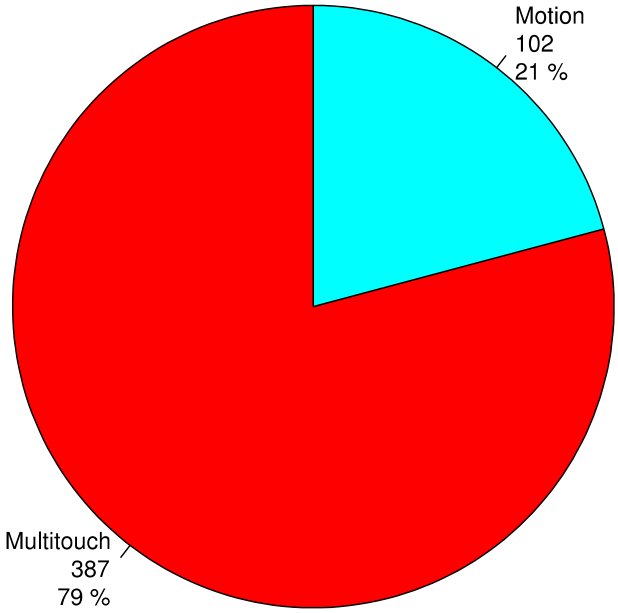


Figure 6.33: Input Modality used to Move the Structure Cursor

Most of the research subjects, namely 79 per cent of them, used multitouch and only 21 per cent of the persons concerned used motion for moving the structure cursor. Hypothesis 1, which proposes that blind and visually impaired persons who are using mobile devices like smart-phones and smart-tablets will choose to use the multitouch input modality for moving the structure cursor, can be confirmed.

The time taken and number of interactions performed by the research subjects for moving the structure cursor from the beginning to the end of the example document in order to obtain a fast overview over the structure of the entire document greatly differs between the cases. Figure 6.34 shows the central tendency, the median and dispersion as well as any outliers covering the entire time and number of interactions range as well as frequencies and percentages of the detailed distribution of the interesting ranges only without the outliers:

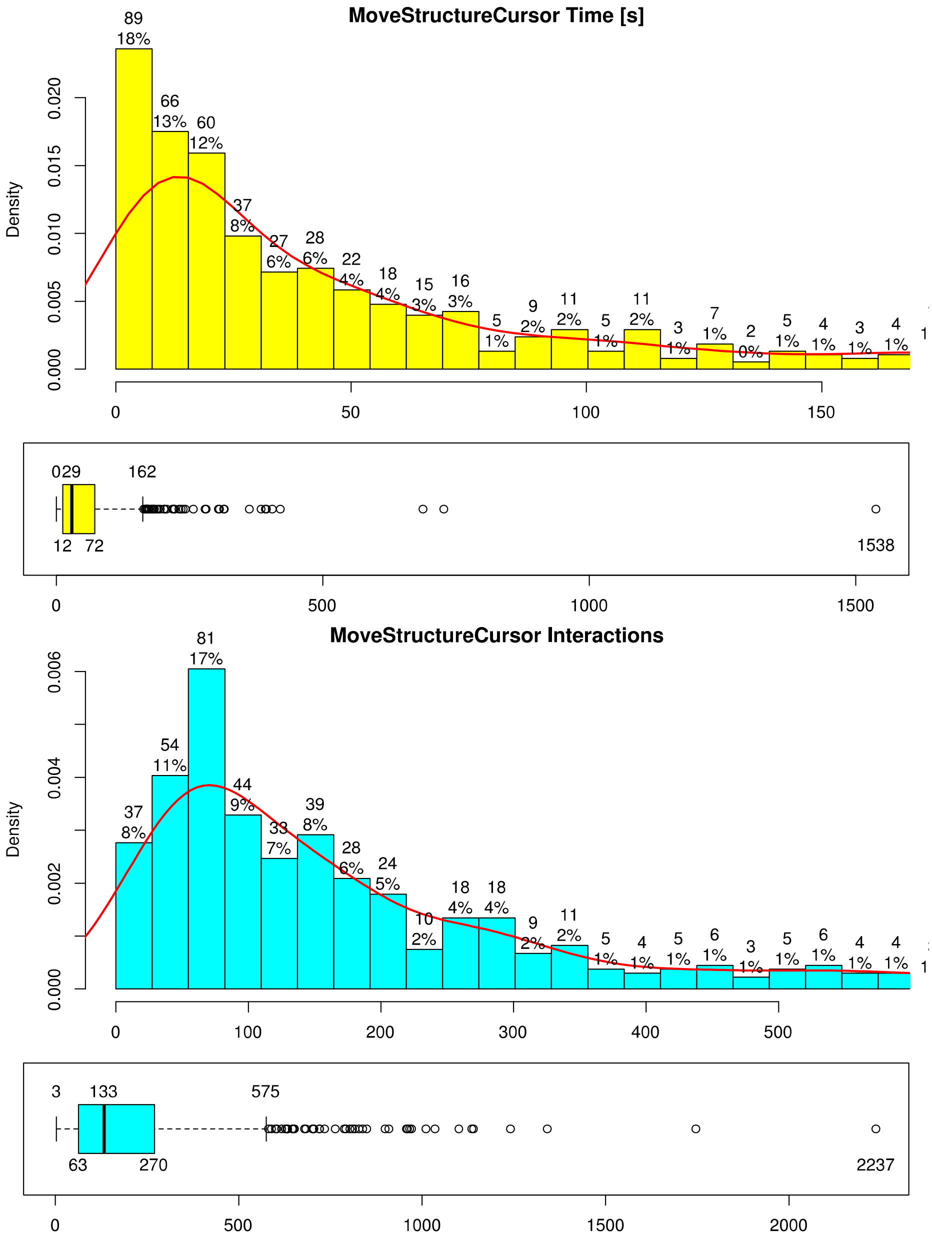


Figure 6.34: Time taken and Interactions performed to Move the Structure Cursor

The middle 50 per cent of research subjects solved the task within 12 to 72 seconds and the median in 29 seconds. The fastest respondents solved the exercise in less than 1 second and the participants slower than any other, other than the outliers who might not participated seriously and took up to 1538 seconds, in 162 seconds. The greatest band of persons concerned within the detailed distribution, namely 18 per cent of them, solved the task within 0 and 8 seconds.

The middle 50 per cent of research subjects solved the task by performing 63 to 270 interactions and the median by performing 133 interactions. The most effective respondents solved the exercise by performing only 3 interactions and the participants less effective than any other, other than the outliers who might not participated seriously and performed up to 2237 interactions, with 575 interactions. The greatest band of persons concerned within the detailed distribution, namely 17 per cent of them, solved the task by performing between 54 and 81 interactions.

This confirms Hypothesis 2, which proposes that the user interface design concepts for moving the structure cursor will be effective and enable blind and visually impaired persons to move the structure cursor forward and backward in order to obtain a fast overview over the structure of an entire document within a usable amount of time taken and with a usable effort of performed interactions.

The effectiveness is dependent on the input modality used by a specific person. The time taken depends to 3 per cent on the input modality, as shown by a weak bivariate correlation with v = 0.18 and a very high level of statistical significance with p = 5.34e-5. The number of interactions performed depends to 2 per cent on the input modality, as shown by a weak bivariate correlation with v = 0.14 and a high level of statistical significance with p = 0.0023. Figure 6.35 shows the central tendency, the median and dispersion as well as any outliers covering the entire time and number of interactions range in correlation to the different input modalities used:

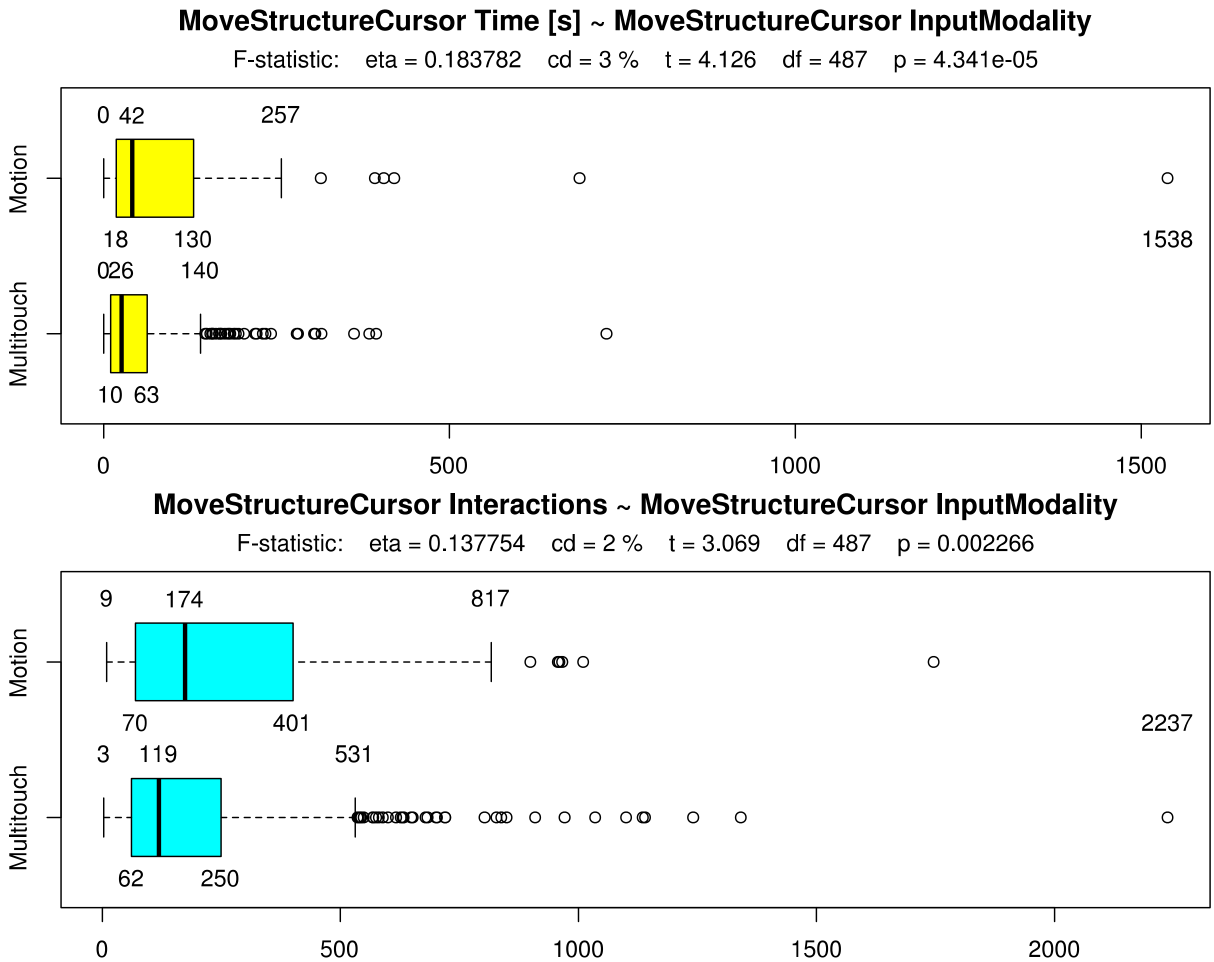


Figure 6.35: Move Structure Cursor Time and Interactions correlated to Input Modality

The middle 50 per cent of research subjects who used multitouch solved the task within 10 to 63 seconds and the median in 26 seconds. The fastest respondents solved the exercise in less than 1 second and the participants slower than any other, other than the outliers, in 140 seconds. On the other hand, the middle 50 per cent of research subjects who used motion solved the task within 18 to 130 seconds and the median in 42 seconds. The fastest respondents solved the exercise in less than 1 second and the participants slower than any other, other than the outliers who took up to 1538 seconds, in 257 seconds.

The middle 50 per cent of research subjects who used multitouch solved the task by performing 62 to 250 interactions and the median by performing 119 interactions. The most effective respondents solved the exercise by performing only 3 interactions and the participants less effective than any other, other than the outliers who performed up to 2237 interactions, with 531 interactions. On the other hand, the middle 50 per cent of research subjects who used motion solved the task by performing 70 to 401 interactions and the median by performing 174 interactions. The most effective respondents solved the exercise by performing only 9 interactions and the participants less effective than any other, other than the outliers, with 817 interactions.

Therefore Hypothesis 3, which proposes that the same effectiveness (time taken and number of interactions performed) to moving the structure cursor will be achieved across all different input modalities, can be rejected.

Hypothesis 4, which proposes that blind and visually impaired persons who are using screen reader will perform better than respondents who are not using screen reader, can be rejected because no statistically significant correlation between the use of screen reader and the time taken and number of interactions performed to moving the structure cursor could be found.

Hypothesis 5, which proposes that the same effectiveness will be achieved across all the different hardware devices employed, can be confirmed because no statistically significant correlation between the hardware device used and the time taken and number of interactions performed to moving the structure cursor could be found.

The number of interactions performed depends to 35 per cent on the time taken to solve the task, as shown by a moderate bivariate correlation with cor = 0.59 and a very high level of statistical significance with p = 0. Figure 6.36 shows the number of interactions performed in correlation to the time taken as well as the linear regression of this relationship:

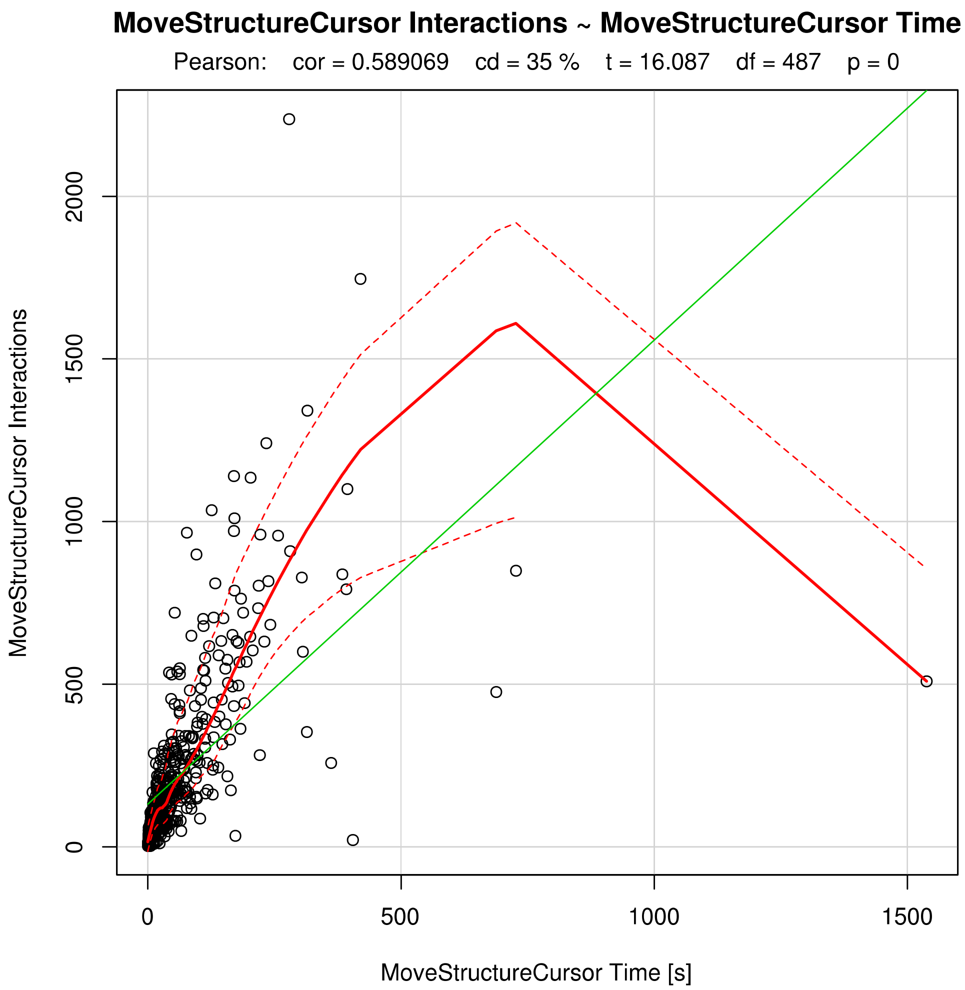


Figure 6.36: Move Structure Cursor Interactions in correlation to Time

This correlation shows that the more time was taken, the more interactions had been performed. These results confirm Hypothesis 6, which proposes that the more effective the user interface design concepts for moving the structure cursor are, the less time will be taken and the less interactions will be performed.

6.4.3 Set Text Cursor

In this exercise, the participants had to set the text cursor to an arbitrary position within a document structure element containing text in order to examine if and how well the respondents are able to setting the text cursor to a position within a document structure element containing text. The text cursor can be set by moving the one pointer more than 50 pixels to the left or to the right or by rotating the device vertically anticlockwise 10 degrees.

The research subjects used different input modalities for setting the text cursor. Figure 6.37 shows the different input modalities used by the persons concerned as well as the frequencies and percentages of respondents using each input modality:

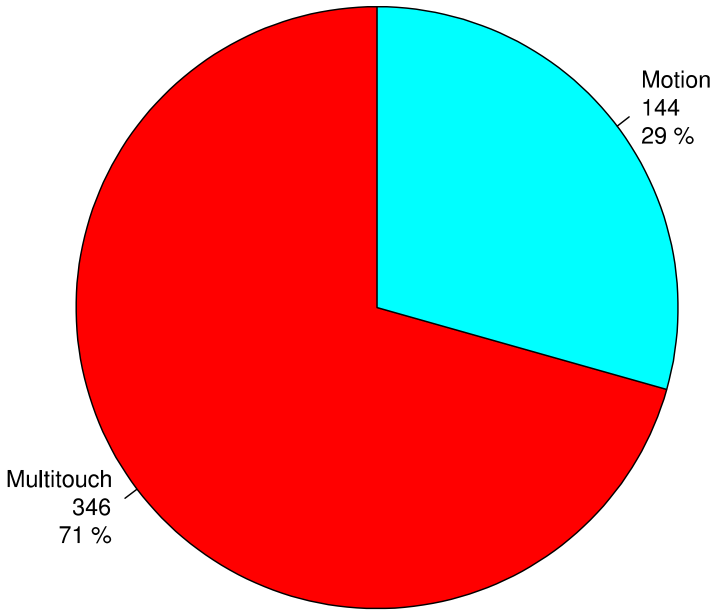


Figure 6.37: Input Modality used to Set the Text Cursor

Most of the research subjects, namely 71 per cent of them, used multitouch and only 29 per cent of the persons concerned used motion for setting the text cursor.

The input modality used depends to 1 per cent on the hardware device used, as shown by a weak bivariate correlation with v = 0.095 and a high level of statistical significance with p = 0.022. Figure 6.38 shows the input modalities used in correlation to the different categories of hardware devices as well as the percentages of how often they are used by each hardware device category:

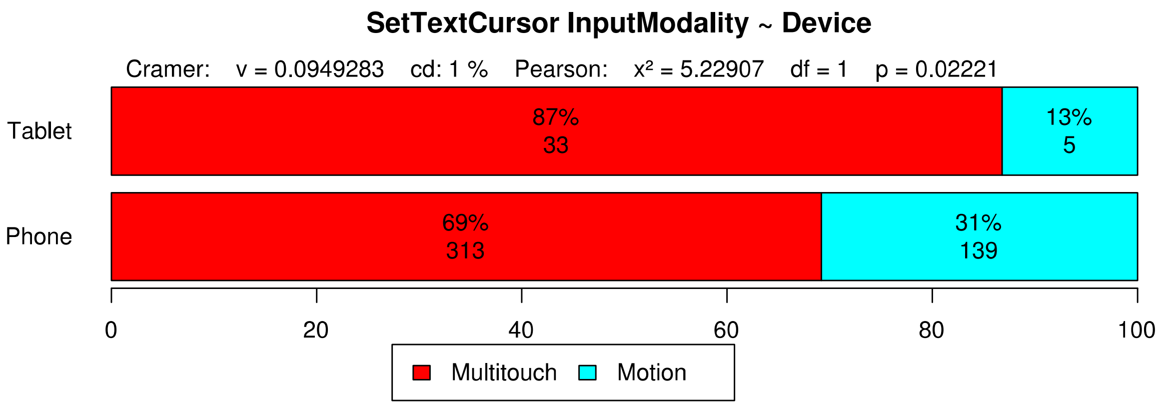


Figure 6.38: Set Text Cursor Input Modality in correlation to Hardware Device

69 per cent of the research subjects using a smart phone as well as 87 per cent of them using a smart tablet employed multitouch. On the other hand, 13 per cent of the respondents using a smart tablet and 31 per cent of them using a smart phone employed motion. Hypothesis 1, which proposes that blind and visually impaired persons who are using mobile devices like smart-phones and smart-tablets will choose to use the multitouch input modality for setting the text cursor, can be confirmed.

The time taken and number of interactions performed by the research subjects for setting the text cursor greatly differs between the cases. Figure 6.39 shows the central tendency, the median and dispersion as well as any outliers covering the entire time and number of interactions range as well as frequencies and percentages of the detailed distribution of the interesting ranges only without the outliers:

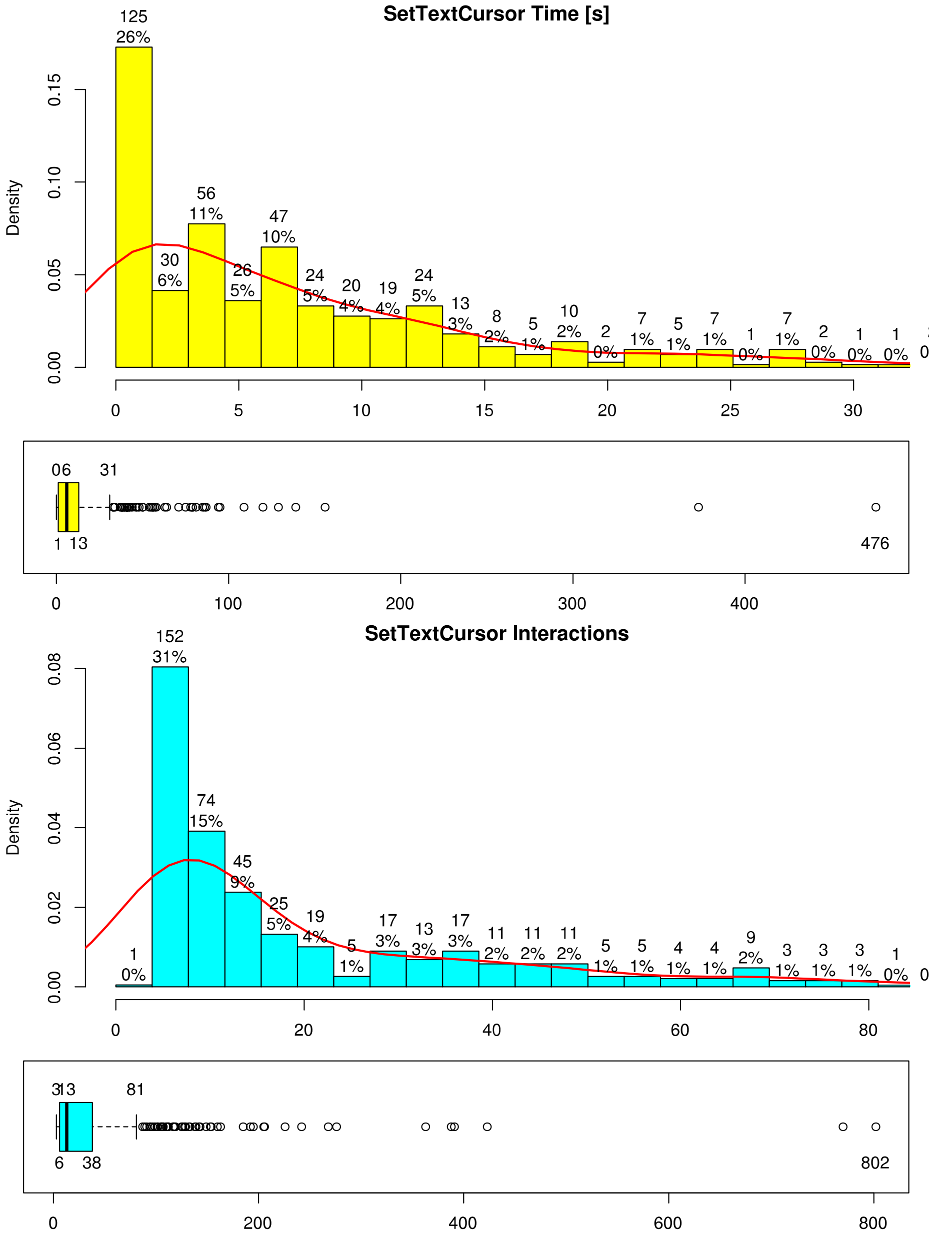


Figure 6.39: Time taken and Interactions performed to Set the Text Cursor

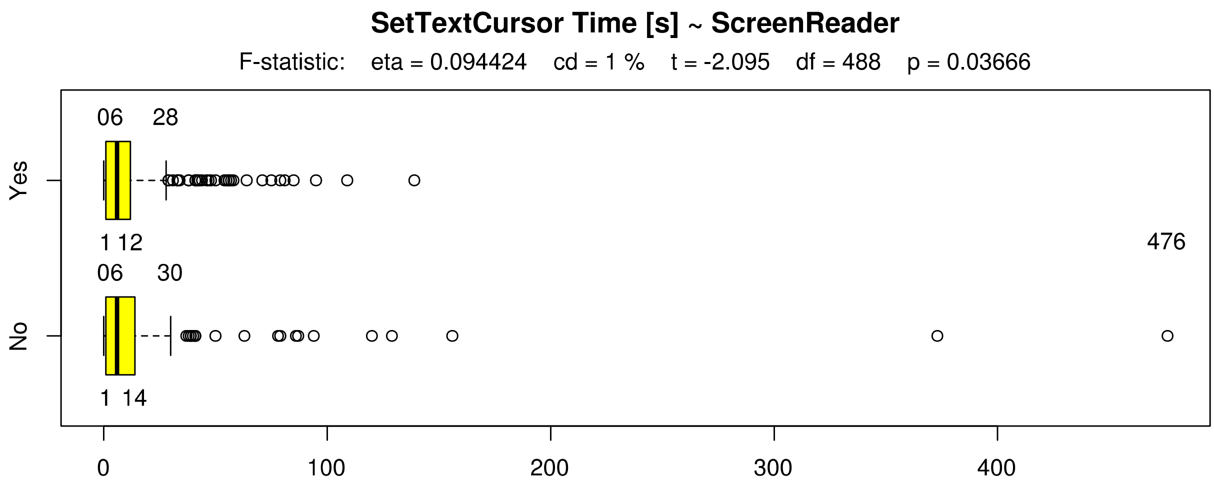
The middle 50 per cent of research subjects solved the task within 1 to 13 seconds and the median in 6 seconds. The fastest respondents solved the exercise in less than 1 second and the participants slower than any other, other than the outliers who might not participated seriously and took up to 476 seconds, in 31 seconds. The greatest band of persons concerned within the detailed distribution, namely 26 per cent of them, solved the task within 0 and 2 seconds.

The middle 50 per cent of research subjects solved the task by performing 6 to 38 interactions and the median by performing 13 interactions. The most effective respondents solved the exercise by performing only 3 interactions and the participants less effective than any other, other than the outliers who might not participated seriously and performed up to 802 interactions, with 81 interactions. The greatest band of persons concerned within the detailed distribution, namely 31 per cent of them, solved the task by performing between 4 and 8 interactions.

This confirms Hypothesis 2, which proposes that the user interface design concepts for setting the text cursor will be effective and enable blind and visually impaired persons to set the text cursor to an arbitrary position within a document structure element containing text within a usable amount of time taken and with a usable effort of performed interactions.

Hypothesis 3, which proposes that the same effectiveness (time taken and number of interactions performed) will be achieved across all different input modalities, can be confirmed because no statistically significant correlation between the input modality used and the time taken and number of interactions performed to setting the text cursor could be found.

The effectiveness is dependent on the use of screen reader by a specific person. The time taken depends to 1 per cent on the use of screen reader, as shown by a weak bivariate correlation with $v = 0.094$ and a high level of statistical significance with $p = 0.037$. The number of interactions performed depends to 1 per cent on the use of screen reader, as shown by a weak bivariate correlation with $v = 0.092$ and a high level of statistical significance with $p = 0.043$. Figure 6.40 shows the central tendency, the median and dispersion as well as any outliers covering the entire time and number of interactions range in correlation to the use of screen reader:



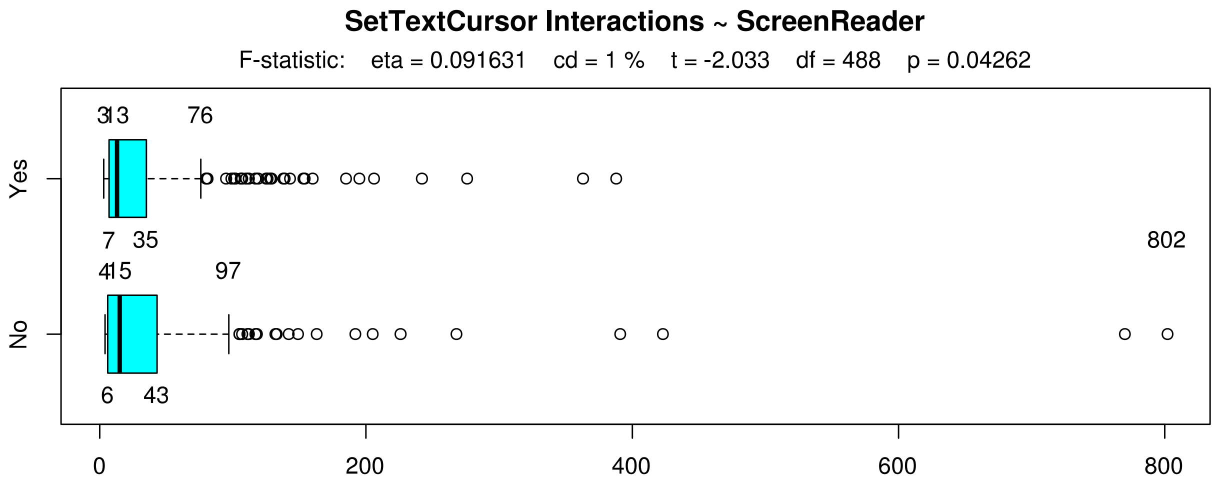


Figure 6.40: Set Text Cursor Time and Interactions correlated to use of Screen Reader

The middle 50 per cent of research subjects who are using screen reader solved the task within 1 to 14 seconds and the median in 6 seconds. The fastest respondents solved the exercise in less than 1 second and the participants slower than any other, other than the outliers, in 28 seconds. On the other hand, the middle 50 per cent of research subjects who are not using screen reader solved the task within 1 to 14 seconds and the median in 6 seconds too. The fastest respondents solved the exercise in less than 1 second and the participants slower than any other, other than the outliers who took up to 476 seconds, in 30 seconds.

The middle 50 per cent of research subjects who are using screen reader solved the task by performing 7 to 35 interactions and the median by performing 13 interactions. The most effective respondents solved the exercise by performing only 3 interactions and the participants less effective than any other, other than the outliers, with 76 interactions. On the other hand, the middle 50 per cent of research subjects who are not using screen reader solved the task by performing 6 to 43 interactions and the median by performing 15 interactions. The most effective respondents solved the exercise by performing only 4 interactions and the participants less effective than any other, other than the outliers who performed up to 802, with 97 interactions.

Therefore Hypothesis 4, which proposes that blind and visually impaired persons who are using screen reader will perform better than respondents who are not using screen reader to setting the text cursor, can be confirmed.

Hypothesis 5, which proposes that the same effectiveness will be achieved across all the different hardware devices employed, can be confirmed because no statistically significant correlation between the hardware device used and the time taken and number of interactions performed to setting the text cursor could be found.

The number of interactions performed depends to 72 per cent on the time taken to solve the task, as shown by a very strong bivariate correlation with cor = 0.85 and a very high level of statistical significance with p = 0. Figure 6.41 shows the number of interactions performed in correlation to the time taken as well as the linear regression of this relationship:

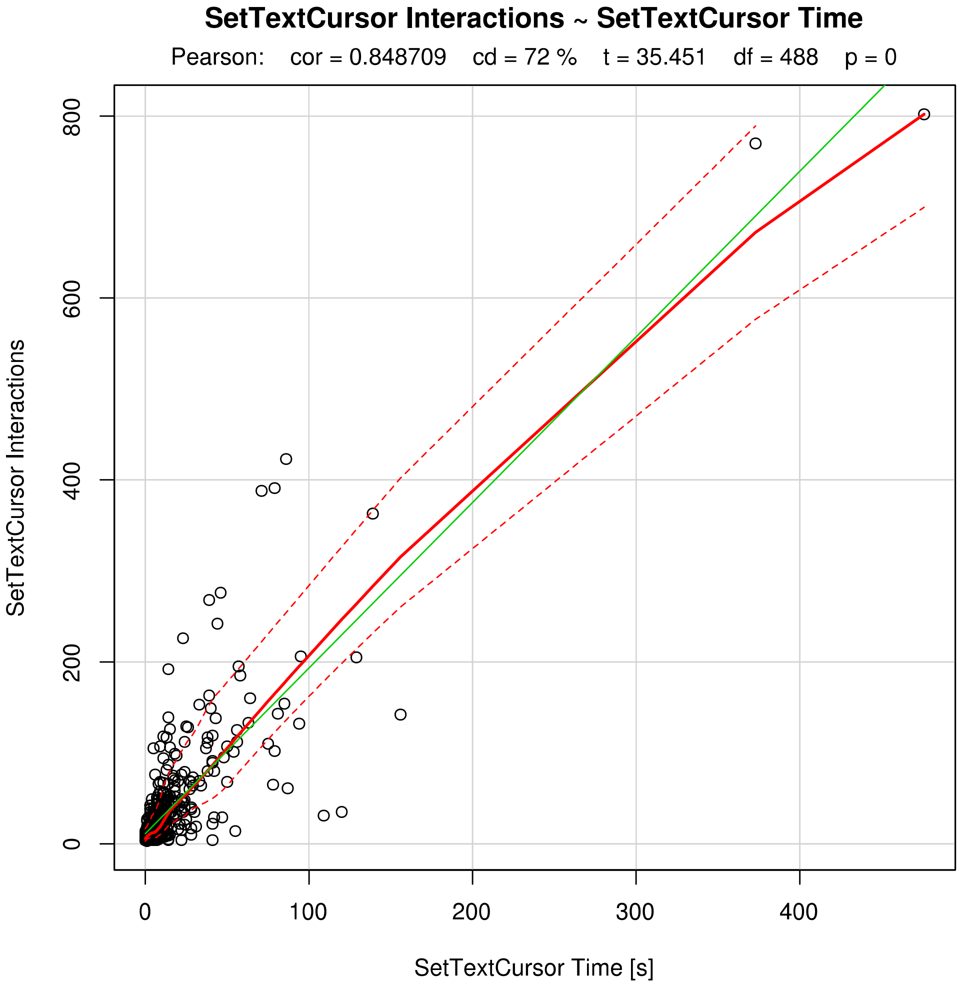


Figure 6.41: Set Text Cursor Interactions in correlation to Time

This correlation shows that the more time was taken, the more interactions had been performed. These results confirm Hypothesis 6, which proposes that the more effective the user interface design concepts for setting the text cursor are, the less time will be taken and the less interactions will be performed.

6.4.4 Move Text Cursor

In this exercise, the participants had to move the text cursor from the beginning to the end of the text contained in an arbitrary document structure element of the example document containing text in order to examine if and how well the respondents are able to moving the text cursor forward and backward in order to obtain a fast overview over the entire text contained in a document structure element. The text cursor can be moved by moving the one pointer left or right at an arbitrary speed or by rotating the device vertically clockwise or anticlockwise within an angle of 90 degrees.

The research subjects used different input modalities for moving the text cursor. Figure 6.42 shows the different input modalities used by the persons concerned as well as the frequencies and percentages of respondents using each input modality:

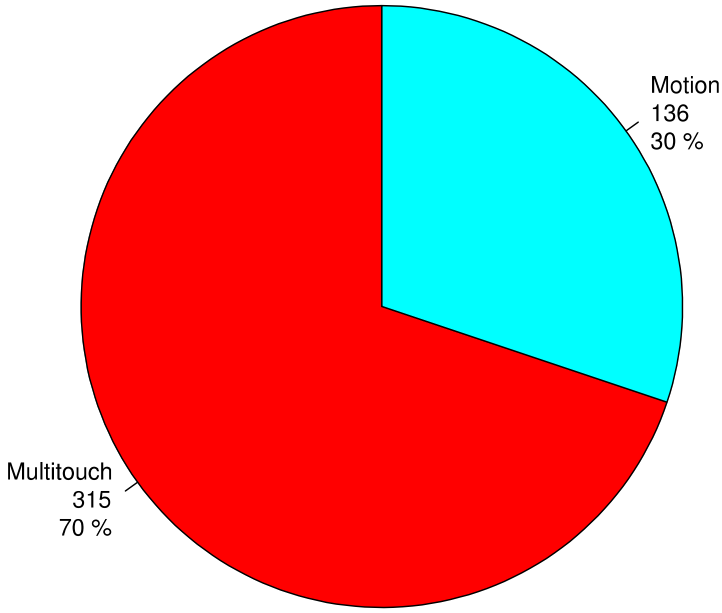


Figure 6.42: Input Modality used to Move the Text Cursor

Most of the research subjects, namely 70 per cent of them, used multitouch and only 30 per cent of the persons concerned used motion for moving the text cursor. Hypothesis 1, which proposes that blind and visually impaired persons who are using mobile devices like smart-phones and smart-tablets will choose to use the multitouch input modality for moving the text cursor, can be confirmed.

The time taken and number of interactions performed by the research subjects for moving the text cursor in order to obtain a fast overview over the entire text contained in a document structure element greatly differs between the cases. Figure 6.43 shows the central tendency, the median and dispersion as well as any outliers covering the entire time and number of interactions range as well as frequencies and percentages of the detailed distribution of the interesting ranges only without the outliers:

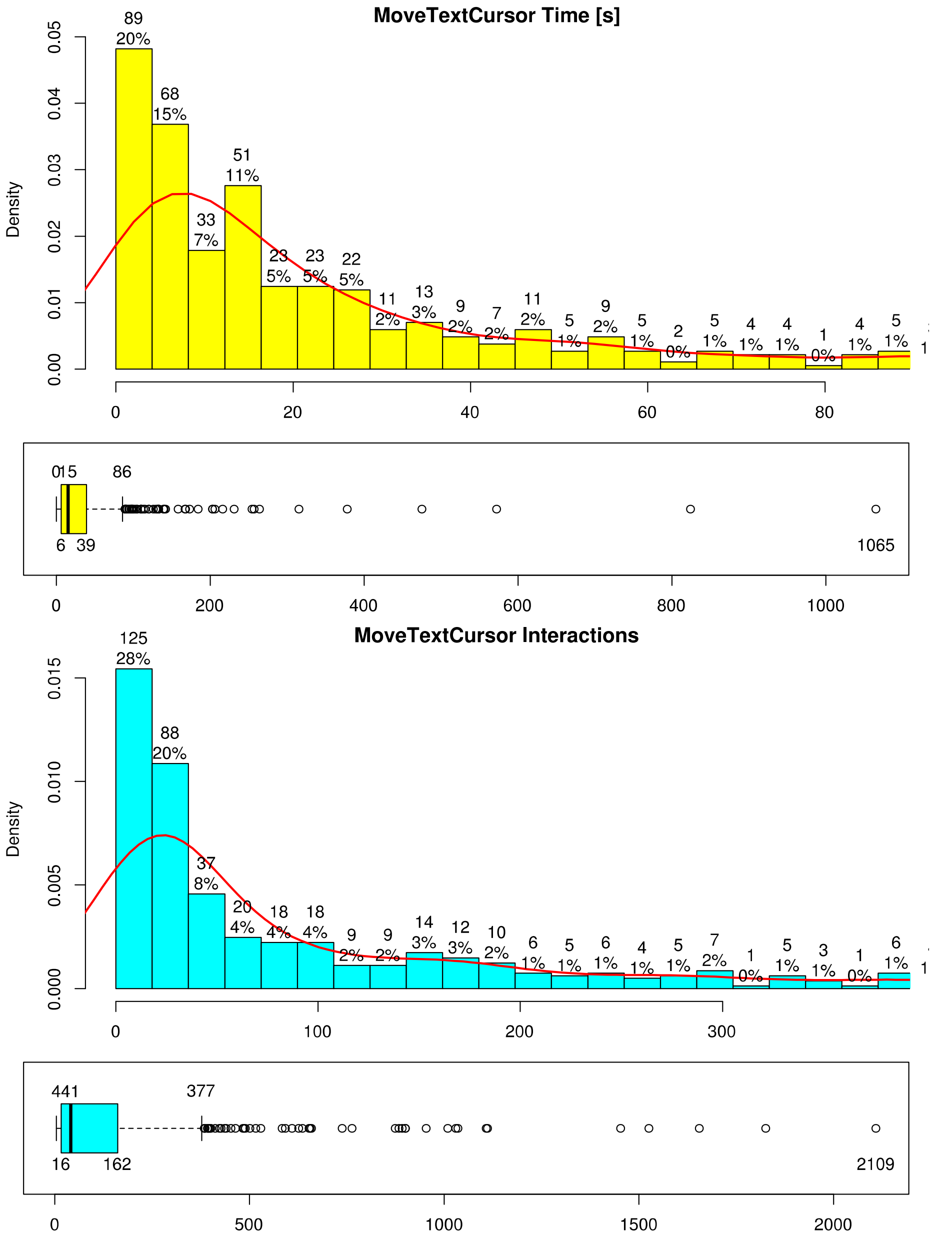


Figure 6.43: Time taken and Interactions performed to Move the Text Cursor

The middle 50 per cent of research subjects solved the task within 6 to 39 seconds and the median in 15 seconds. The fastest respondents solved the exercise in less than 1 second and the participants slower than any other, other than the outliers who might not participated seriously and took up to 1065 seconds, in 86 seconds. The greatest band of persons concerned within the detailed distribution, namely 20 per cent of them, solved the task within 0 and 4 seconds.

The middle 50 per cent of research subjects solved the task by performing 16 to 162 interactions and the median by performing 41 interactions. The most effective respondents solved the exercise by performing only 4 interactions and the participants less effective than any other, other than the outliers who might not participated seriously and performed up to 2109 interactions, with 377 interactions. The greatest band of persons concerned within the detailed distribution, namely 28 per cent of them, solved the task by performing between 0 and 18 interactions.

This confirms Hypothesis 2, which proposes that the user interface design concepts for moving the text cursor will be effective and enable blind and visually impaired persons to to move the text cursor forward and backward in order to obtain a fast overview over the entire text contained in a document structure element within a usable amount of time taken and with a usable effort of performed interactions.

The effectiveness is dependent on the input modality used by a specific person. The time taken depends to 4 per cent on the input modality, as shown by a weak bivariate correlation with v = 0.19 and a very high level of statistical significance with p = 5.78e-5. The number of interactions performed depends to 5 per cent on the input modality, as shown by another weak bivariate correlation with v = 0.23 and a very high level of statistical significance with p = 5.83e-7. Figure 6.44 shows the central tendency, the median and dispersion as well as any outliers covering the entire time and number of interactions range in correlation to the different input modalities used:

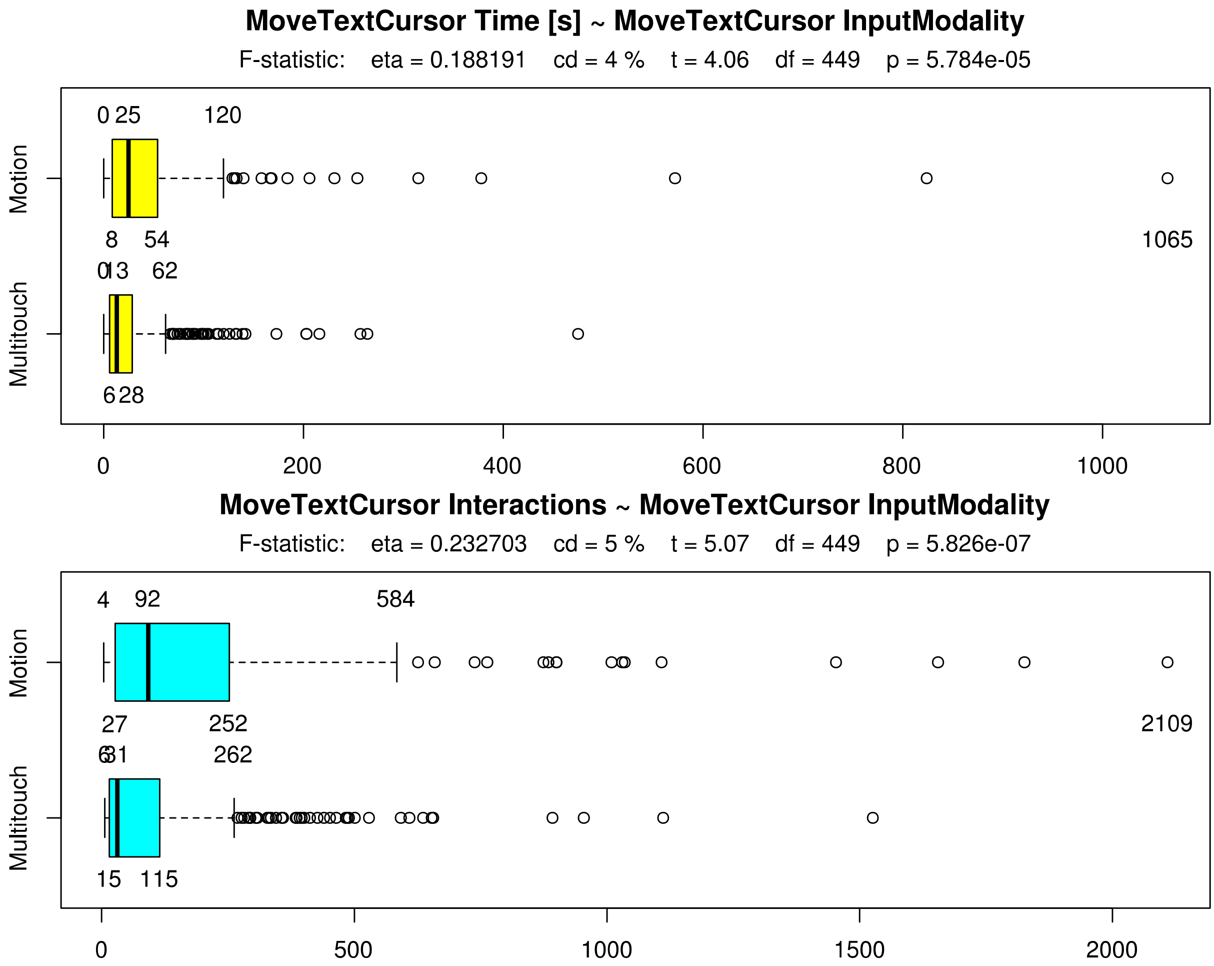


Figure 6.44: Move Text Cursor Time and Interactions in correlation to Input Modality

The middle 50 per cent of research subjects who used multitouch solved the task within 6 to 28 seconds and the median in 13 seconds. The fastest respondents solved the exercise in less than 1 second and the participants slower than any other, other than the outliers, in 62 seconds. On the other hand, the middle 50 per cent of research subjects who used motion solved the task within 8 to 54 seconds and the median in 25 seconds. The fastest respondents solved the exercise in less than 1 second and the participants slower than any other, other than the outliers who took up to 1065 seconds, in 120 seconds.

The middle 50 per cent of research subjects who used multitouch solved the task by performing 15 to 115 interactions and the median by performing 31 interactions. The most effective respondents solved the exercise by performing only 6 interactions and the participants less effective than any other, other than the outliers, with 262 interactions. On the other hand, the middle 50 per cent of research subjects who used motion solved the task by performing 27 to 252 interactions and the median by performing 92 interactions. The most effective respondents solved the exercise by performing only 4 interactions and the participants less effective than any other, other than the outliers who performed up to 2109, with 584 interactions.

Therefore Hypothesis 3, which proposes that the same effectiveness (time taken and number of interactions performed) to moving the text cursor will be achieved across all different input modalities, can be rejected.

Hypothesis 4, which proposes that blind and visually impaired persons who are using screen reader will perform better than respondents who are not using screen reader, can be rejected because no statistically significant correlation between the use of screen reader and the time taken and number of interactions performed to moving the text cursor could be found.

Hypothesis 5, which proposes that the same effectiveness will be achieved across all the different hardware devices employed, can be confirmed because no statistically significant correlation between the hardware device used and the time taken and number of interactions performed to moving the text cursor could be found.

The number of interactions performed depends to 62 per cent on the time taken to solve the task, as shown by a strong bivariate correlation with cor = 0.79 and a very high level of statistical significance with $p = 0$. Figure 6.45 shows the number of interactions performed in correlation to the time taken as well as the linear regression of this relationship:

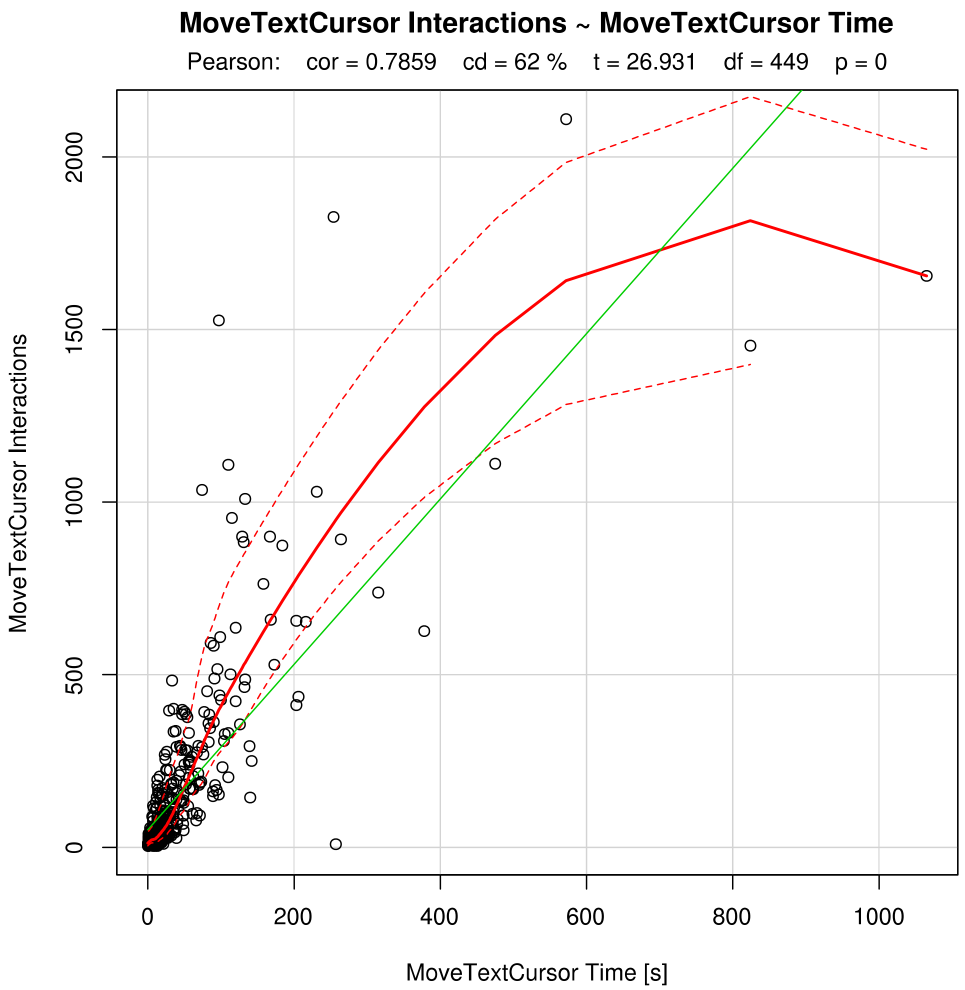


Figure 6.45: Move Text Cursor Interactions in correlation to Time

This correlation shows that the more time was taken, the more interactions had been performed. These results confirm Hypothesis 6, which proposes that the more effective the user interface design concepts for moving the text cursor are, the less time will be taken and the less interactions will be performed.

6.4.5 Unset Text Cursor

In this exercise, the participants had to unset the text cursor at an arbitrary position within a document structure element containing text in order to examine if and how well the respondents are able to unsetting the text cursor at a position within a document structure element containing text. The text cursor can be unset by moving the one pointer more than 50 pixels up or down or by rotating the device vertically back to the start position.

The research subjects used different input modalities for unsetting the text cursor. Figure 6.46 shows the different input modalities used by the persons concerned as well as the frequencies and percentages of respondents using each input modality:

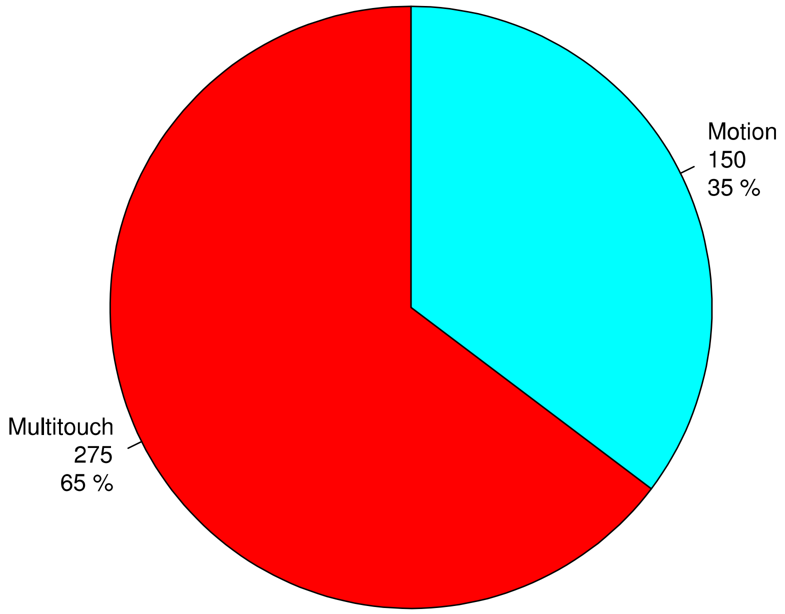


Figure 6.46: Input Modality used to Unset the Text Cursor

Most of the research subjects, namely 65 per cent of them, used multitouch and only 35 per cent of the persons concerned used motion for unsetting the text cursor.

The input modality used depends to 1 per cent on the hardware device used, as shown by a weak bivariate correlation with v = 0.090 and a high level of statistical significance with p = 0.042. Figure 6.47 shows the input modalities used in correlation to the different categories of hardware devices as well as the percentages of how often they are used by each hardware device category:

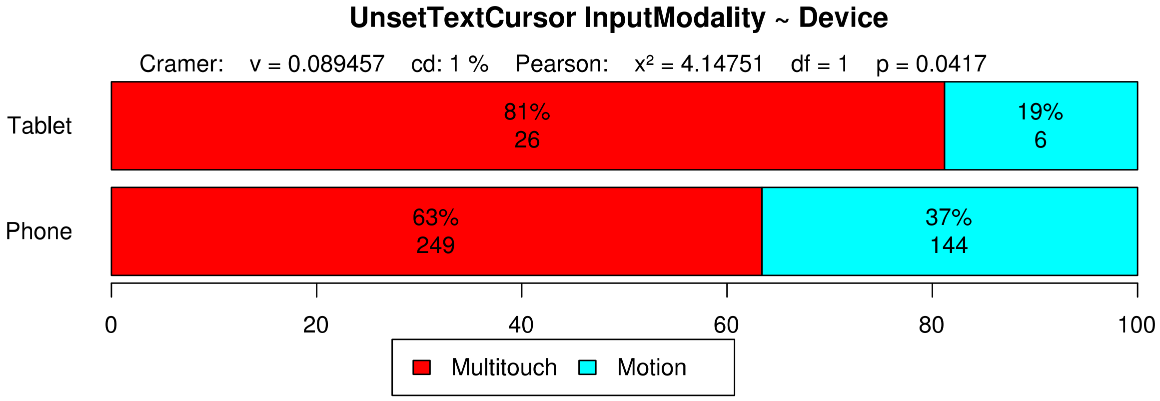


Figure 6.47: Unset Text Cursor Input Modality in correlation to Hardware Device

63 per cent of the research subjects using a smart phone as well as 81 per cent of them using a smart tablet employed multitouch. On the other hand, only 19 per cent of the respondents using a smart tablet and 37 per cent of them using a smart phone employed motion. Hypothesis 1, which proposes that blind and visually impaired persons who are using mobile devices like smart-phones and smart-tablets will choose to use the multitouch input modality for unsetting the text cursor, can be confirmed.

The time taken and number of interactions performed by the research subjects for unsetting the text cursor greatly differs between the cases. Figure 6.48 shows the central tendency, the median and dispersion as well as any outliers covering the entire time and number of interactions range as well as frequencies and percentages of the detailed distribution of the interesting ranges only without the outliers:

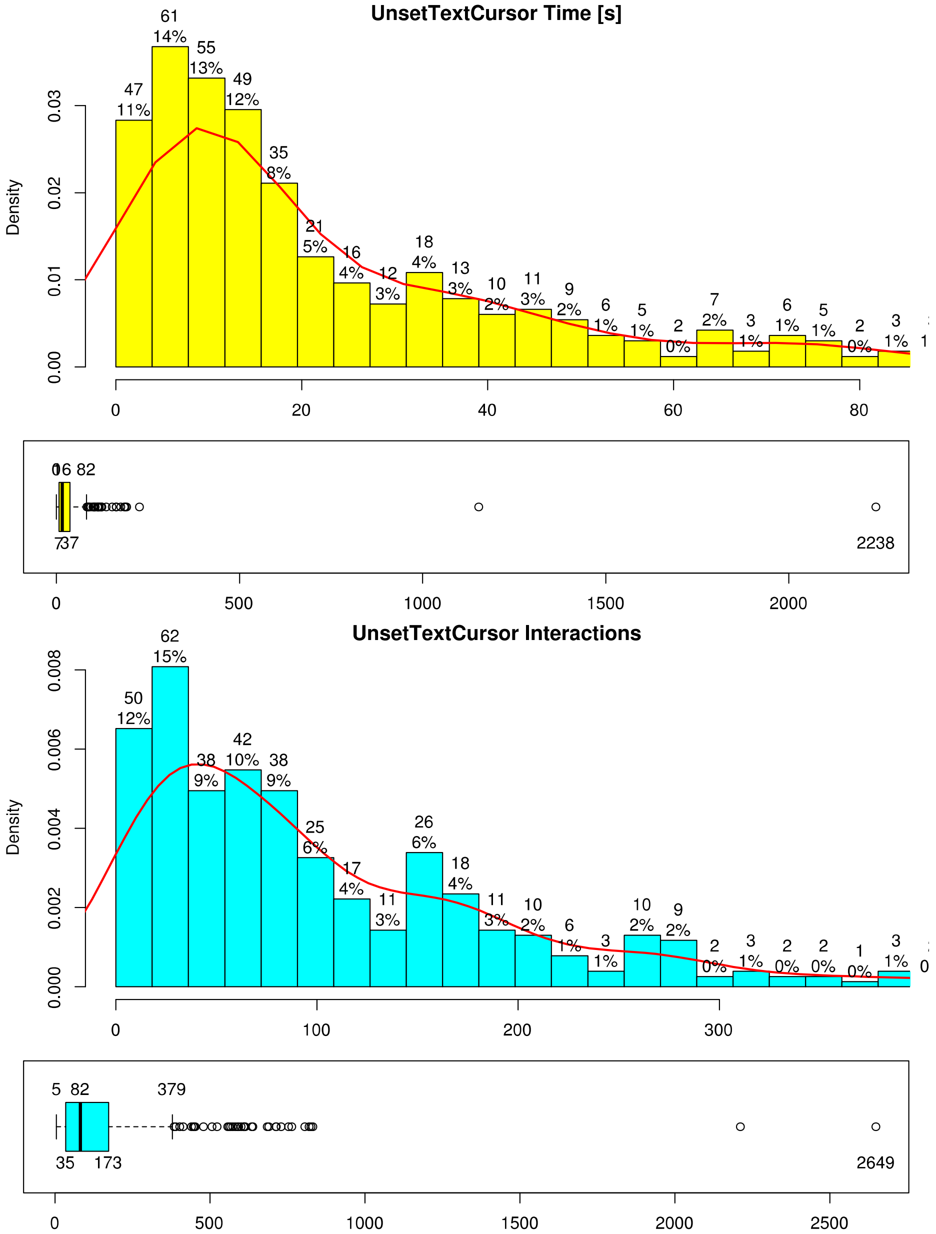


Figure 6.48: Time taken and Interactions performed to Unset the Text Cursor

The middle 50 per cent of research subjects solved the task within 7 to 37 seconds and the median in 16 seconds. The fastest respondents solved the exercise in less than 1 second and the participants slower than any other, other than the outliers who might not participated seriously and took up to 2238 seconds, in 82 seconds. The greatest band of persons concerned within the detailed distribution, namely 14 per cent of them, solved the task within 4 and 8 seconds.

The middle 50 per cent of research subjects solved the task by performing 35 to 173 interactions and the median by performing 82 interactions. The most effective respondents solved the exercise by performing only 5 interactions and the participants less effective than any other, other than the outliers who might not participated seriously and performed up to 2649 interactions, with 379 interactions. The greatest band of persons concerned within the detailed distribution, namely 15 per cent of them, solved the task by performing between 18 and 36 interactions.

This confirms Hypothesis 2, which proposes that the user interface design concepts for unsetting the text cursor will be effective and enable blind and visually impaired persons to unset the text cursor at an arbitrary position within a document structure element containing text within a usable amount of time taken and with a usable effort of performed interactions.

Hypothesis 3, which proposes that the same effectiveness (time taken and number of interactions performed) will be achieved across all different input modalities, can be confirmed because no statistically significant correlation between the hardware input modality used and the time taken and number of interactions performed to moving the structure cursor could be found.

Hypothesis 4, which proposes that blind and visually impaired persons who are using screen reader will perform better than respondents who are not using screen reader, can be rejected because no statistically significant correlation between the use of screen reader and the time taken and number of interactions performed to unsetting the text cursor could be found.

The effectiveness is dependent on the hardware device used by a specific person. The number of interactions performed depends to 1 per cent on the hardware device, as shown by a weak bivariate correlation with v = 0.097 and a high level of statistical significance with p = 0.047. Figure 6.49 shows the central tendency, the median and dispersion as well as any outliers covering the entire number of interactions range in correlation to the different hardware devices used:

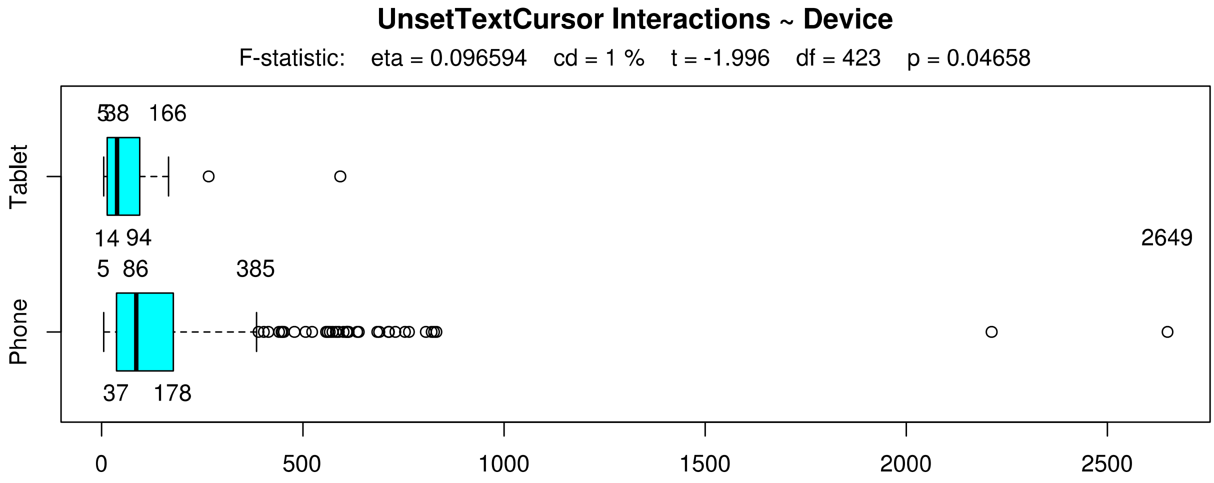


Figure 6.49: Unset Text Cursor Interactions in correlation to Hardware Device

The middle 50 per cent of research subjects who used a smart tablet solved the task by performing 14 to 97 interactions and the median by performing 38 interactions. The most effective respondents solved the exercise by performing only 5 interactions and the participants less effective than any other, other than the outliers, with 166 interactions.

On the other hand, the middle 50 per cent of research subjects who used a smart phone solved the task by performing 37 to 178 interactions and the median by performing 86 interactions. The most effective respondents solved the exercise by performing only 5 interactions and the participants less effective than any other, other than the outliers who performed up to 2649 interactions, with 385 interactions.

Therefore Hypothesis 5, which proposes that the same effectiveness will be achieved across all the different hardware devices employed to unsetting the text cursor, can be rejected.

The number of interactions performed depends to 59 per cent on the time taken to solve the task, as shown by a strong bivariate correlation with cor = 0.77 and a very high level of statistical significance with p = 0. Figure 6.50 shows the number of interactions performed in correlation to the time taken as well as the linear regression of this relationship:

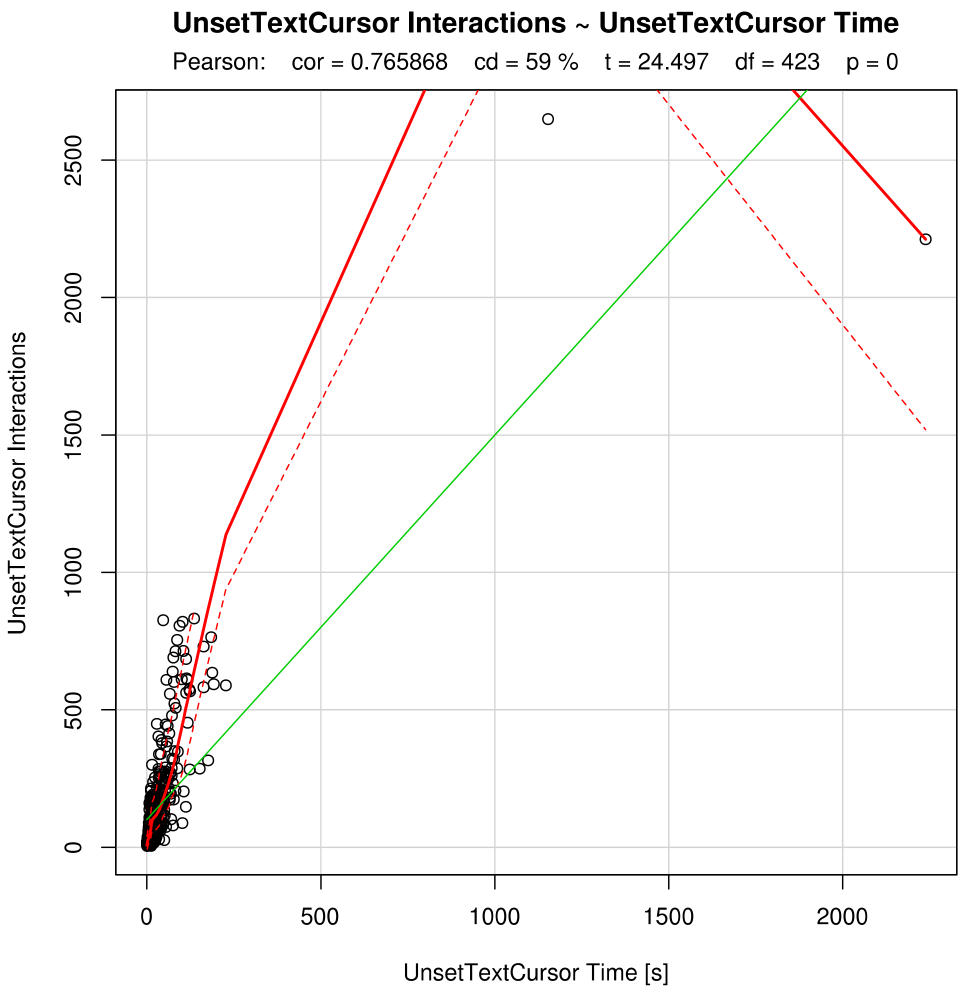


Figure 6.50: Unset Text Cursor Interactions in correlation to Time

This correlation shows that the more time was taken, the more interactions had been performed. These results confirm Hypothesis 6, which proposes that the more effective the user interface design concepts for unsetting the text cursor are, the less time will be taken and the less interactions will be performed.

6.4.6 Unset Structure Cursor

In this exercise, the participants had to unset the structure cursor at an arbitrary position within the document in order to examine if and how well the respondents are able to unsetting the structure cursor at a position within the document. The structure cursor can be unset by lifting up the one pointer or by rotating the device horizontally back to the start position.

The research subjects used different input modalities for unsetting the structure cursor. Figure 6.51 shows the different input modalities used by the persons concerned as well as the frequencies and percentages of respondents using each input modality:

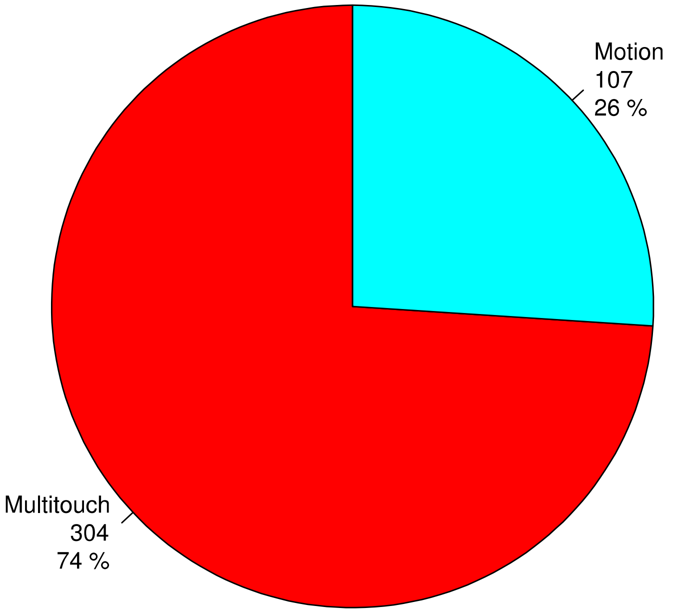


Figure 6.51: Input Modality used to Unset the Structure Cursor

Most of the research subjects, namely 74 per cent of them, used multitouch and only 26 per cent of the persons concerned used motion for unsetting the structure cursor. Hypothesis 1, which proposes that blind and visually impaired persons who are using mobile devices like smart-phones and smart-tablets will choose to use the multitouch input modality for unsetting the structure cursor, can be confirmed.

The time taken and number of interactions performed by the research subjects for unsetting the structure cursor greatly differs between the cases. Figure 6.52 shows the central tendency, the median and dispersion as well as any outliers covering the entire time and number of interactions range as well as frequencies and percentages of the detailed distribution of the interesting ranges only without the outliers:

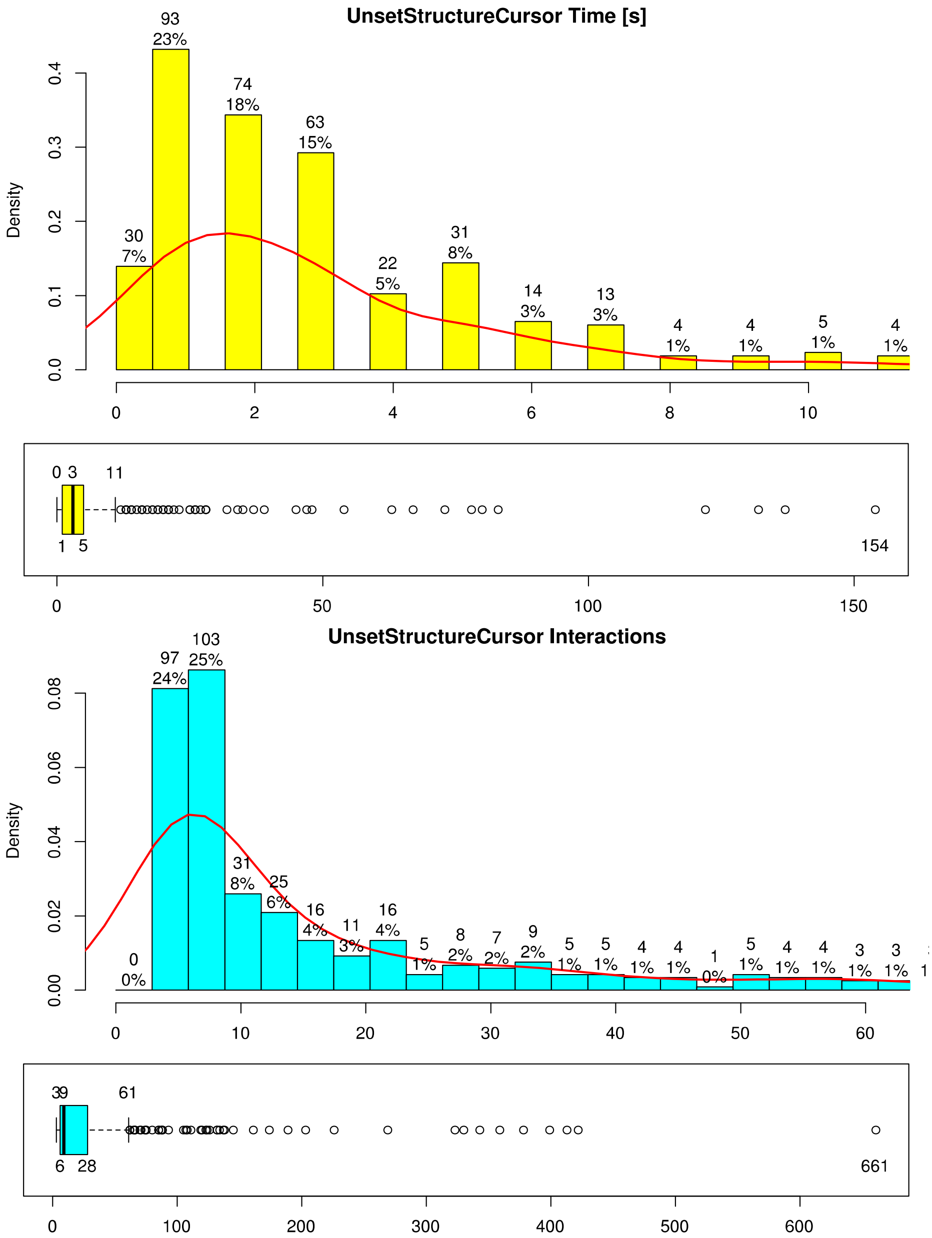


Figure 6.52: Time taken and Interactions performed to Unset the Structure Cursor

The middle 50 per cent of research subjects solved the task within 1 to 5 seconds and the median in 3 seconds. The fastest respondents solved the exercise in less than 1 second and the participants slower than any other, other than the outliers who might not participated seriously and took up to 154 seconds, in 11 seconds. The greatest band of persons concerned within the detailed distribution, namely 23 per cent of them, solved the task within 1 second.

The middle 50 per cent of research subjects solved the task by performing 6 to 28 interactions and the median by performing 9 interactions. The most effective respondents solved the exercise by performing only 3 interactions and the participants less effective than any other, other than the outliers who might not participated seriously and performed up to 661 interactions, with 61 interactions. The greatest band of persons concerned within the detailed distribution, namely 25 per cent of them, solved the task by performing between 6 and 9 interactions.

This confirms Hypothesis 2, which proposes that the user interface design concepts for unsetting the structure cursor will be effective and enable blind and visually impaired persons to unset the text cursor at an arbitrary position within the document within a usable amount of time taken and with a usable effort of performed interactions.

The effectiveness is dependent on the input modality used by a specific person. The time taken depends to 13 per cent on the input modality, as shown by a weak bivariate correlation with v = 0.36 and a very high level of statistical significance with p = 4.04e-14. The number of interactions performed depends to 19 per cent on the input modality, as shown by a moderate bivariate correlation with v = 0.44 and a very high level of statistical significance with p = 8.51e-21. Figure 6.53 shows the central tendency, the median and dispersion as well as any outliers covering the entire time and number of interactions range in correlation to the different input modalities used:

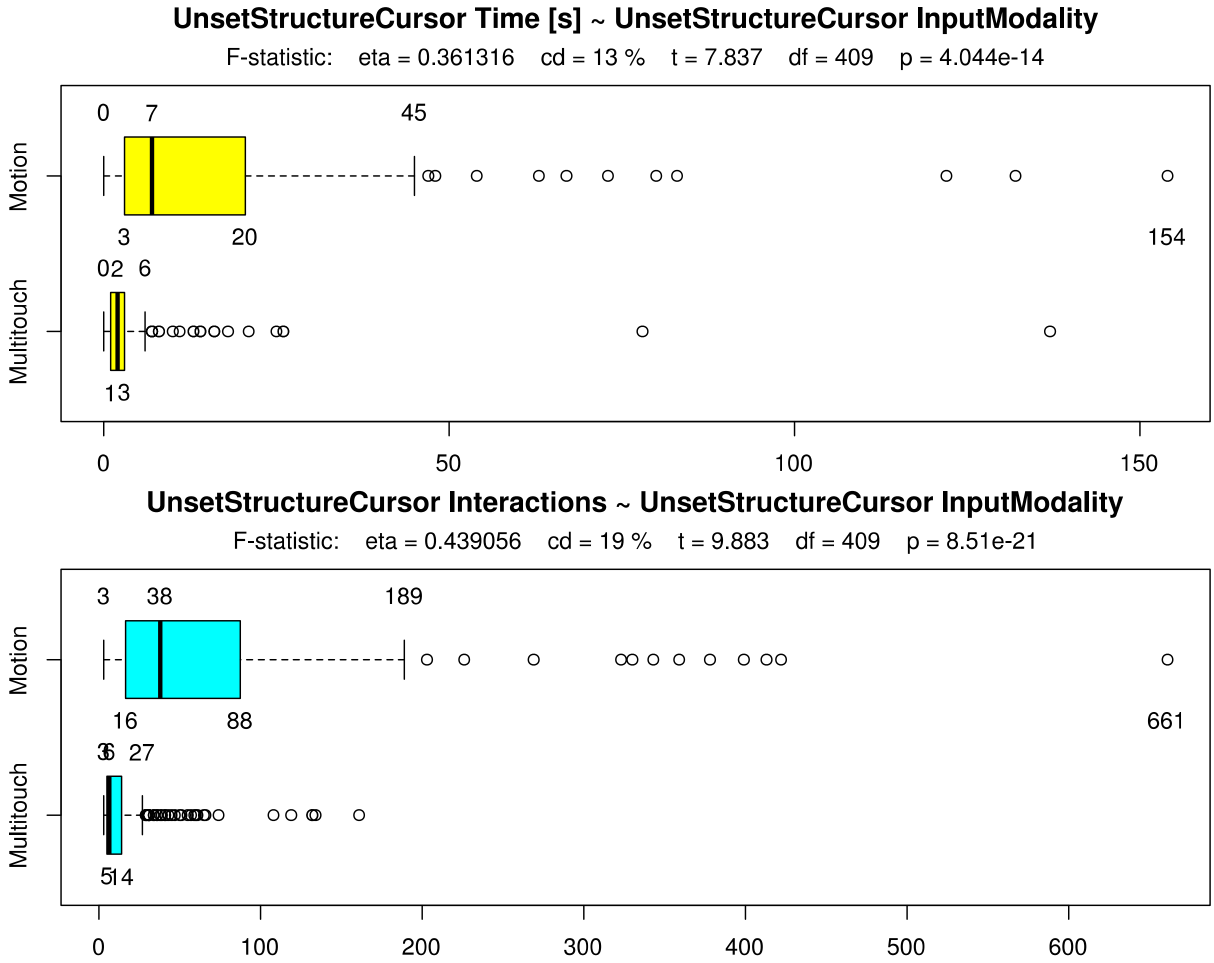


Figure 6.53: Unset Structure Cursor Time and Interactions correlated to Input Modality

The middle 50 per cent of research subjects who used multitouch solved the task within 1 to 3 seconds and the median in 2 seconds. The fastest respondents solved the exercise in less than 1 second and the participants slower than any other, other than the outliers, in 6 seconds. On the other hand, the middle 50 per cent of research subjects who used motion solved the task within 3 to 20 seconds and the median in 7 seconds. The fastest respondents solved the exercise in less than 1 second and the participants slower than any other, other than the outliers who took up to 154 seconds, in 45 seconds.

The middle 50 per cent of research subjects who used multitouch solved the task by performing 5 to 14 interactions and the median by performing 6 interactions. The most effective respondents solved the exercise by performing only 3 interactions and the participants less effective than any other, other than the outliers, with 27 interactions. On the other hand, the middle 50 per cent of research subjects who used motion solved the task by performing 16 to 88 interactions and the median by performing 38 interactions. The most effective respondents solved the exercise by performing only 3 interactions and the participants less effective than any other, other than the outliers who performed up to 661, with 189 interactions.

Therefore Hypothesis 3, which proposes that the same effectiveness (time taken and number of interactions performed) to unsetting the structure cursor will be achieved across all different input modalities, can be rejected.

Hypothesis 4, which proposes that blind and visually impaired persons who are using screen reader will perform better than respondents who are not using screen reader, can be rejected because no statistically significant correlation between the use of screen reader and the time taken and number of interactions performed to unsetting the structure cursor could be found.

Hypothesis 5, which proposes that the same effectiveness will be achieved across all the different hardware devices employed, can be confirmed because no statistically significant correlation between the hardware device used and the time taken and number of interactions performed to unsetting the structure cursor could be found.

The number of interactions performed depends to 64 per cent on the time taken to solve the task, as shown by a very strong bivariate correlation with cor = 0.80 and a very high level of statistical significance with p = 0. Figure 6.54 shows the number of interactions performed in correlation to the time taken as well as the linear regression of this relationship:

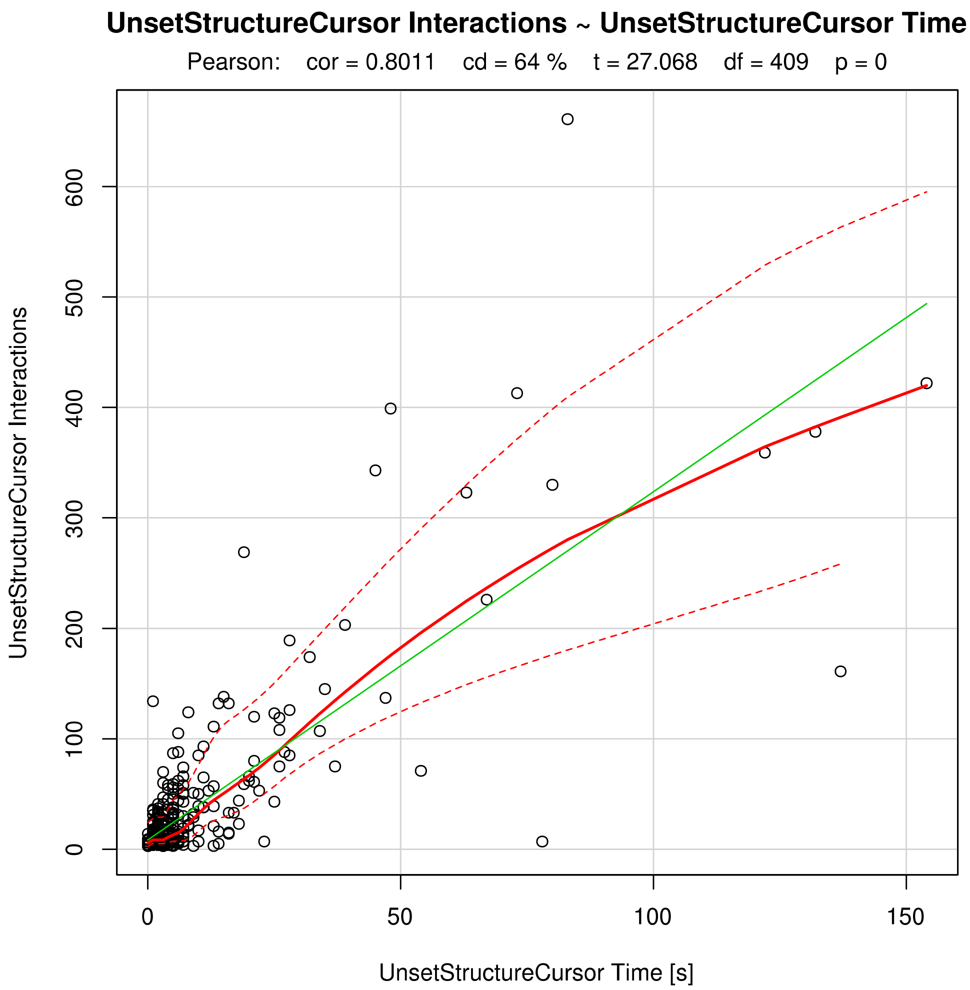


Figure 6.54: Unset Structure Cursor Interactions in correlation to Time

This correlation shows that the more time was taken, the more interactions had been performed. These results confirm Hypothesis 6, which proposes that the more effective the user interface design concepts for unsetting the structure cursor are, the less time will be taken and the less interactions will be performed.

6.4.7 Set Select Modifier

In this exercise, the participants had to select an arbitrary document structure element or text node in order to examine if and how well the respondents are able to setting the select modifier. The select modifier can be set by putting down an additional second pointer at an arbitrary position or by flipping the device 90 degrees to the left.

The research subjects used different input modalities for setting the select modifier. Figure 6.55 shows the different input modalities used by the persons concerned as well as the frequencies and percentages of respondents using each input modality:

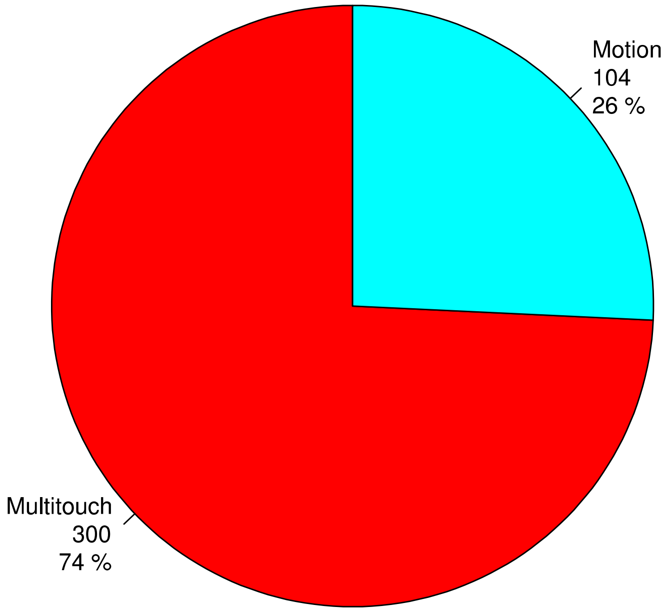


Figure 6.55: Input Modality used to Set the Select Modifier

Most of the research subjects, namely 74 per cent of them, used multitouch and only 26 per cent of the persons concerned used motion for setting the select modifier.

The input modality used depends to 1 per cent on the hardware device used, as shown by a weak bivariate correlation with v = 0.095 and a high level of statistical significance with p = 0.033. Figure 6.56 shows the input modalities used in correlation to the different categories of hardware devices as well as the percentages of how often they are used by each hardware device category:

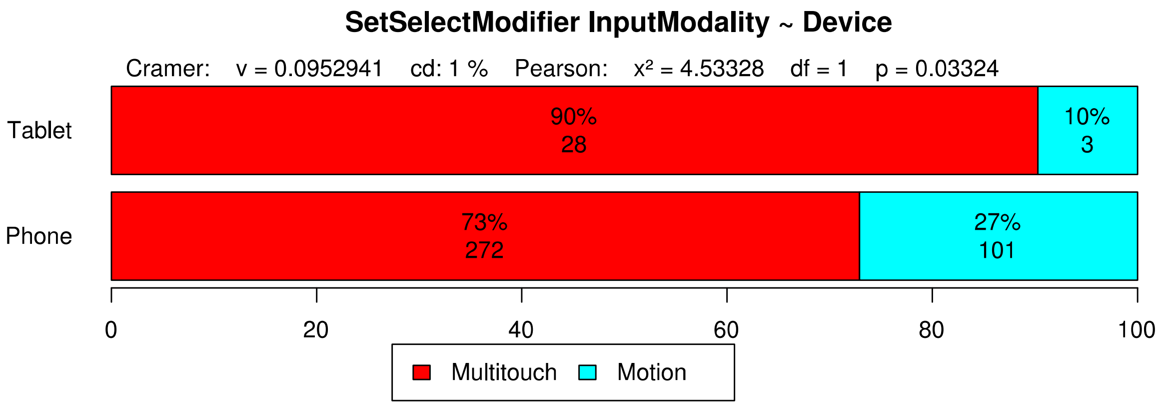


Figure 6.56: Set Select Modifier Input Modality in correlation to Hardware Device

73 per cent of the research subjects using a smart phone as well as 90 per cent of them using a smart tablet employed multitouch. On the other hand, only 10 per cent of the

respondents using a smart tablet and 27 per cent of them using a smart phone employed motion. Hypothesis 1, which proposes that blind and visually impaired persons who are using mobile devices like smart-phones and smart-tablets will choose to use the multitouch input modality for setting the select modifier, can be confirmed.

The time taken and number of interactions performed by the research subjects for setting the select modifier greatly differs between the cases. Figure 6.57 shows the central tendency, the median and dispersion as well as any outliers covering the entire time and number of interactions range as well as frequencies and percentages of the detailed distribution of the interesting ranges only without the outliers:

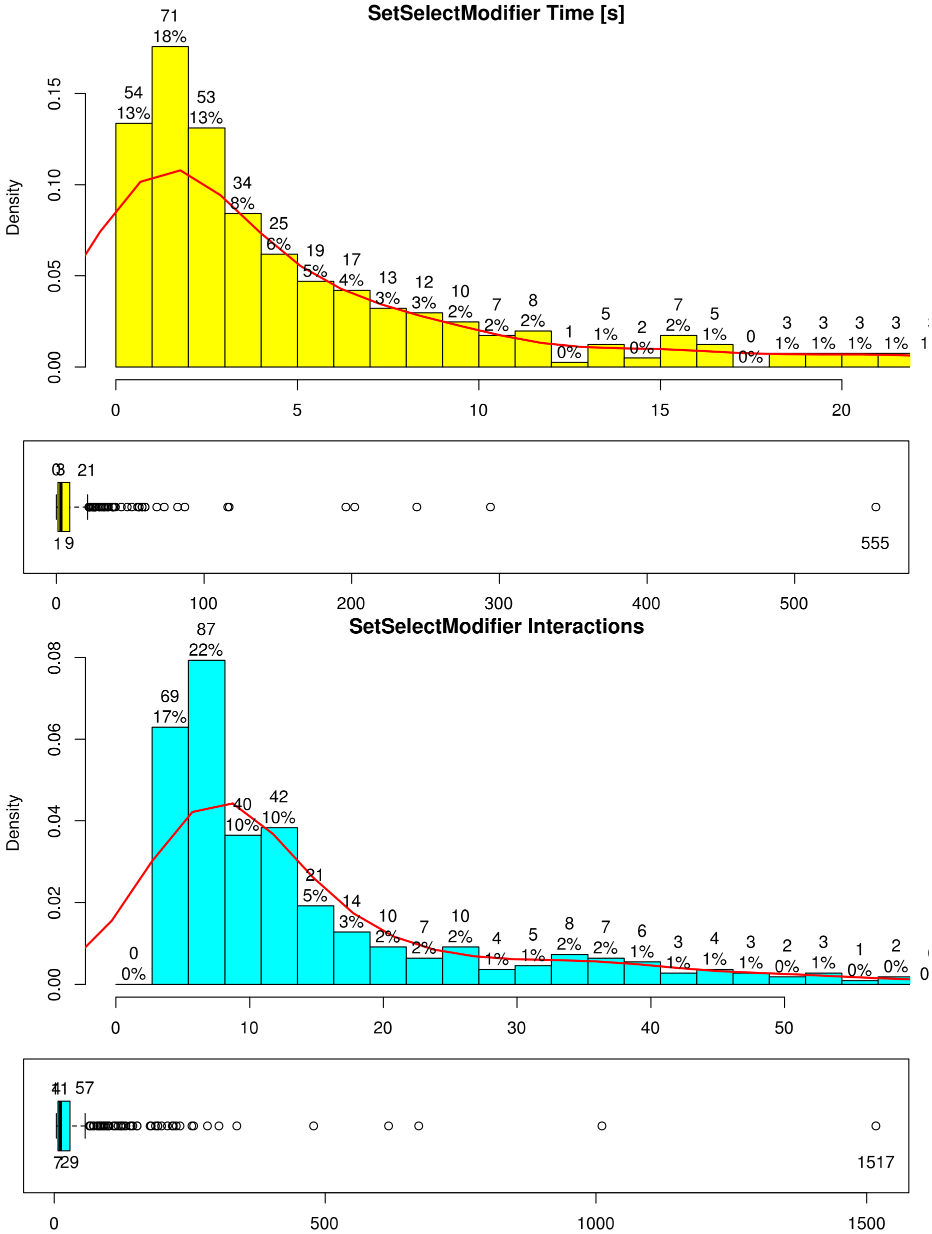


Figure 6.57: Time taken and Interactions performed to Set the Select Modifier

The middle 50 per cent of research subjects solved the task within 1 to 9 seconds and the median in 8 seconds. The fastest respondents solved the exercise in less than 1 second and the participants slower than any other, other than the outliers who might not participated seriously and took up to 555 seconds, in 21 seconds. The greatest band of persons concerned within the detailed distribution, namely 18 per cent of them, solved the task within 2 seconds.

The middle 50 per cent of research subjects solved the task by performing 7 to 29 interactions and the median by performing 11 interactions. The most effective respondents solved the exercise by performing only 4 interactions and the participants less effective than any other, other than the outliers who might not participated seriously and performed up to 1517 interactions, with 57 interactions. The greatest band of persons concerned within the detailed distribution, namely 22 per cent of them, solved the task by performing between 6 and 9 interactions.

This confirms Hypothesis 2, which proposes that the user interface design concepts for setting the select modifier will be effective and enable blind and visually impaired persons to select a document structure element or text node within a usable amount of time taken and with a usable effort of performed interactions.

The effectiveness is dependent on the input modality used by a specific person. The time taken depends to 2 per cent on the input modality, as shown by a weak bivariate correlation with v = 0.15 and a high level of statistical significance with p = 0.0021. The number of interactions performed depends to 2 per cent on the input modality, as shown by another weak bivariate correlation with v = 0.13 and a high level of statistical significance with p = 0.0089. Figure 6.58 shows the central tendency, the median and dispersion as well as any outliers covering the entire time and number of interactions range in correlation to the different input modalities used:

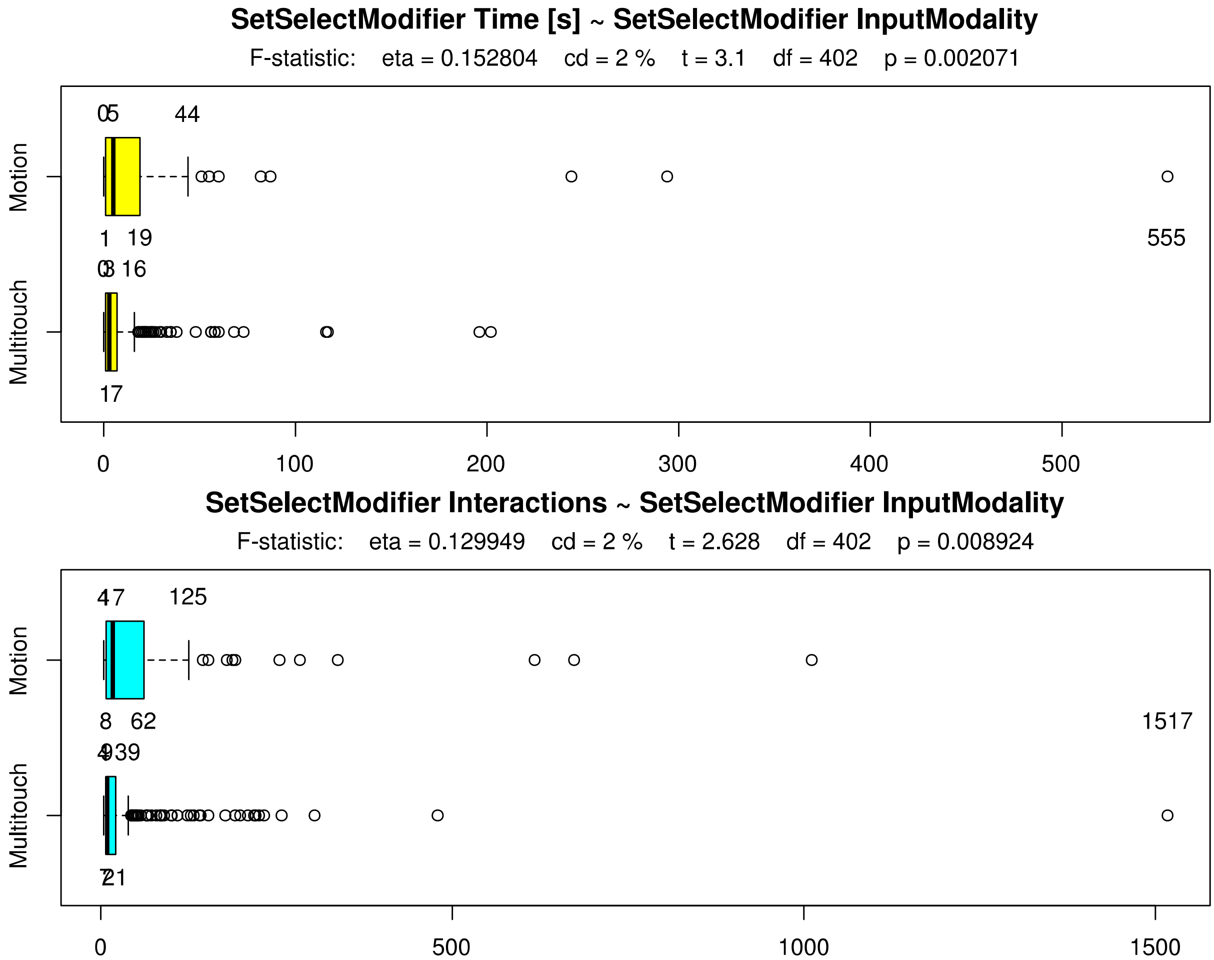


Figure 6.58: Set Select Modifier Time and Interactions in correlation to Input Modality

The middle 50 per cent of research subjects who used multitouch solved the task within 1 to 7 seconds and the median in 3 seconds. The fastest respondents solved the exercise in less than 1 second and the participants slower than any other, other than the outliers, in 16 seconds. On the other hand, the middle 50 per cent of research subjects who used motion solved the task within 1 to 19 seconds and the median in 5 seconds. The fastest respondents solved the exercise in less than 1 second and the participants slower than any other, other than the outliers who took up to 555 seconds, in 44 seconds.

The middle 50 per cent of research subjects who used multitouch solved the task by performing 7 to 21 interactions and the median by performing 9 interactions. The most effective respondents solved the exercise by performing only 4 interactions and the participants less effective than any other, other than the outliers who performed up to 1517 interactions, with 39 interactions. On the other hand, the middle 50 per cent of research subjects who used motion solved the task by performing 8 to 62 interactions and the median by performing 7 interactions. The most effective respondents solved the exercise by performing only 4 interactions and the participants less effective than any other, other than the outliers, with 125 interactions.

Therefore Hypothesis 3, which proposes that the same effectiveness (time taken and number of interactions performed) to setting the select modifier will be achieved across all different input modalities, can be rejected.

Hypothesis 4, which proposes that blind and visually impaired persons who are using screen reader will perform better than respondents who are not using screen reader, can be rejected because no statistically significant correlation between the use of screen reader and the time taken and number of interactions performed to setting the select modifier could be found.

Hypothesis 5, which proposes that the same effectiveness will be achieved across all the different hardware devices employed, can be confirmed because no statistically significant correlation between the hardware device used and the time taken and number of interactions performed to setting the select modifier could be found.

The number of interactions performed depends to 64 per cent on the time taken to solve the task, as shown by a strong bivariate correlation with cor = 0.80 and a very high level of statistical significance with $p = 0$. Figure 6.59 shows the number of interactions performed in correlation to the time taken as well as the linear regression of this relationship:

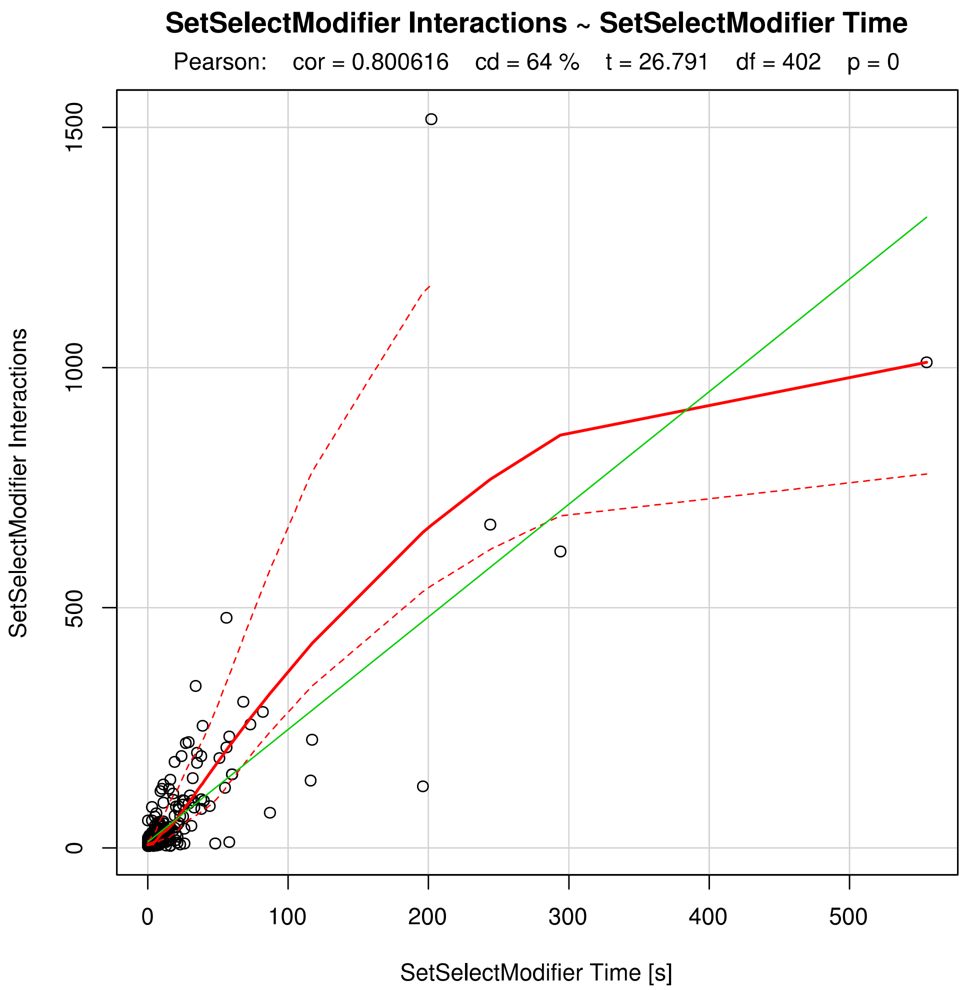


Figure 6.59: Set Select Modifier Interactions in correlation to Time

This correlation shows that the more time was taken, the more interactions had been performed. These results confirm Hypothesis 6, which proposes that the more effective the user interface design concepts for setting the select modifier are, the less time will be taken and the less interactions will be performed.

6.5 Manipulation

The last 6 exercises focussing on the user interaction design concepts and their hypotheses about the behaviour of blind and visually impaired persons among structured documents in relation to the interactions for manipulation provided by the user interaction design as described in Chapter 5.3.

6.5.1 Activate Selection

In this exercise, the participants had to activate an arbitrary selection of the document for manipulation in order to examine if and how well the respondents are able to activate a selection of the document for manipulation. When a selection of the document is selected, it can be activated for manipulation by setting the active modifier by using one of the different input modalities.

The research subjects used different input modalities for activating a selection. Figure 6.60 shows the different input modalities used by the persons concerned as well as the frequencies and percentages of respondents using each input modality:

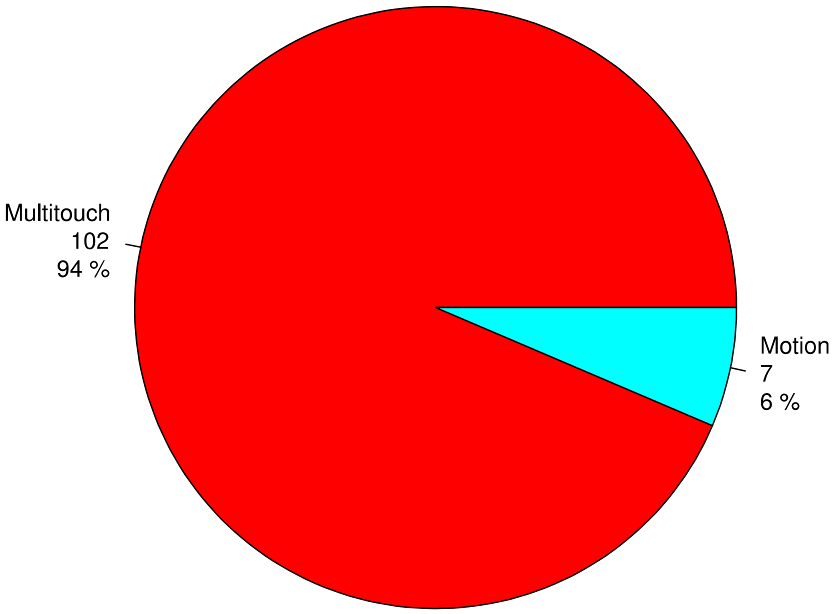


Figure 6.60: Input Modality used to Activate a Selection

Most of the research subjects, namely 94 per cent of them, used multitouch and only 6 per cent of the persons concerned used motion for activating a selection. Hypothesis 1, which proposes that blind and visually impaired persons who are using mobile devices like smart-phones and smart-tablets will choose to use the multitouch input modality for activating a selection, can be confirmed.

The time taken and number of interactions performed by the research subjects for activating a selection greatly differs between the cases. Figure 6.61 shows the central tendency, the median and dispersion as well as any outliers covering the entire time and number of interactions range as well as frequencies and percentages of the detailed distribution of the interesting ranges only without the outliers:

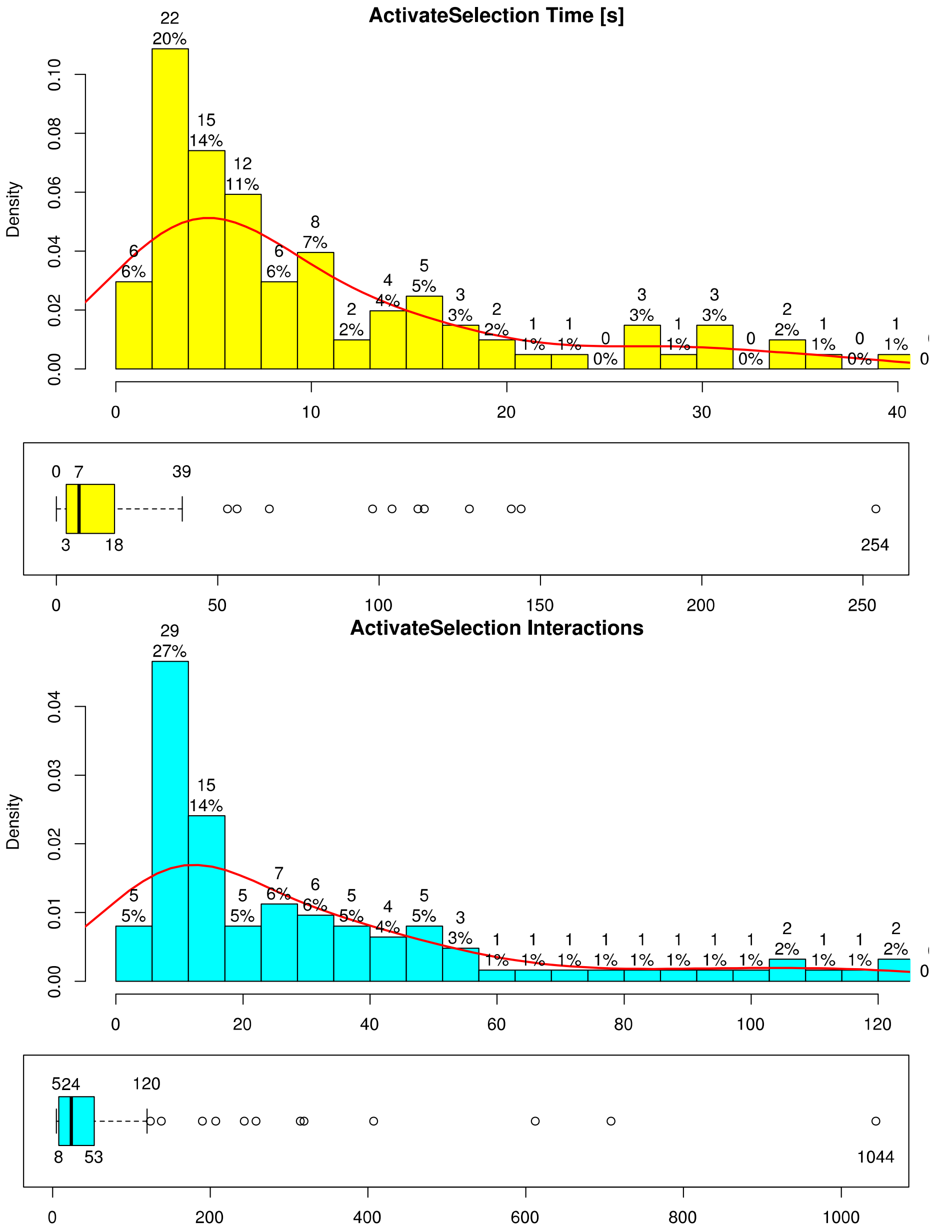


Figure 6.61: Time taken and Interactions performed to Activate a Selection

The middle 50 per cent of research subjects solved the task within 3 to 18 seconds and the median in 7 seconds. The fastest respondents solved the exercise in less than 1 second and the participants slower than any other, other than the outliers who might not participated seriously and took up to 254 seconds, in 39 seconds. The greatest band of persons concerned within the detailed distribution, namely 18 per cent of them, solved the task within 2 and 4 seconds.

The middle 50 per cent of research subjects solved the task by performing 8 to 53 interactions and the median by performing 24 interactions. The most effective respondents solved the exercise by performing only 5 interactions and the participants less effective than any other, other than the outliers who might not participated seriously and performed up to 1044 interactions, with 120 interactions. The greatest band of persons concerned within the detailed distribution, namely 27 per cent of them, solved the task by performing between 6 and 12 interactions.

This confirms Hypothesis 2, which proposes that the user interaction design concept for activating a selection will be effective and enable blind and visually impaired persons to activate a selection of the document for manipulation within a usable amount of time taken and with a usable effort of performed interactions.

The effectiveness is dependent on the input modality used by a specific person. The time taken depends to 6 per cent on the input modality, as shown by a weak bivariate correlation with eta = 0.25 and a very high level of statistical significance with p = 0.0087. The number of interactions performed depends to 17 per cent on the input modality, as shown by a moderate bivariate correlation with eta = 0.41 and very a high level of statistical significance with p = 1.12e-5. Figure 6.62 shows the central tendency, the median and dispersion as well as any outliers covering the entire time and number of interactions range in correlation to the different input modalities used:

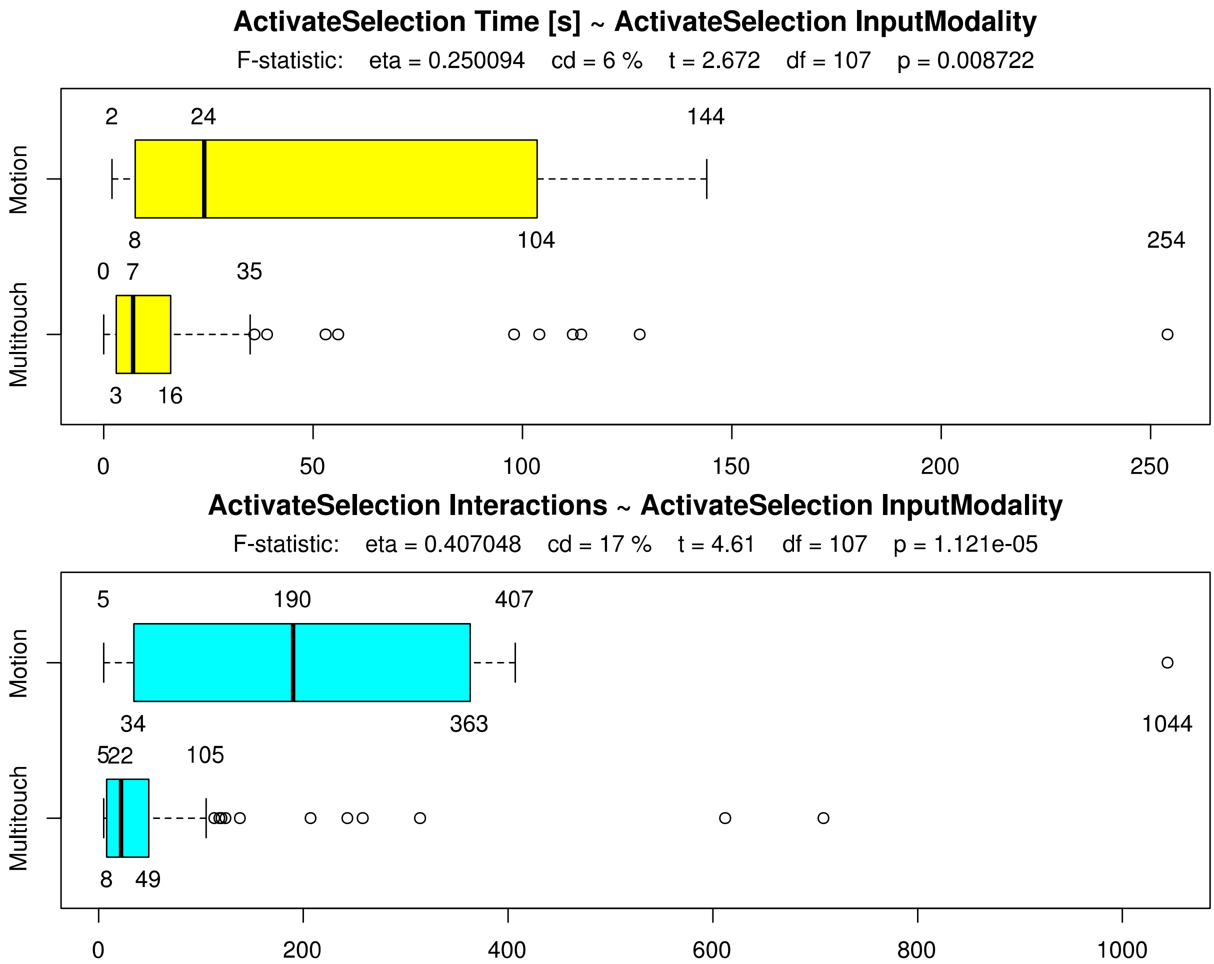


Figure 6.62: Activate Selection Time and Interactions in correlation to Input Modality

The middle 50 per cent of research subjects who used multitouch solved the task within 3 to 16 seconds and the median in 7 seconds. The fastest respondents solved the exercise in less than 1 second and the participants slower than any other, other than the outliers who took up to 254 seconds, in 35 seconds. On the other hand, the middle 50 per cent of research subjects who used motion solved the task within 8 to 104 seconds and the median in 24 seconds. The fastest respondents solved the exercise in 2 seconds and the participants slower than any other, other than the outliers, in 144 seconds.

The middle 50 per cent of research subjects who used multitouch solved the task by performing 8 to 49 interactions and the median by performing 22 interactions. The most effective respondents solved the exercise by performing only 5 interactions and the participants less effective than any other, other than the outliers, with 105 interactions. On the other hand, the middle 50 per cent of research subjects who used motion solved

the task by performing 34 to 363 interactions and the median by performing 190 interactions. The most effective respondents solved the exercise by performing only 5 interactions and the participants less effective than any other, other than the outliers who performed up to 1044 interactions, with 407 interactions.

Therefore Hypothesis 3, which proposes that the same effectiveness (time taken and number of interactions performed) to activating a selection will be achieved across all different input modalities, can be rejected.

Hypothesis 4, which proposes that blind and visually impaired persons who are using screen reader will perform better than respondents who are not using screen reader, can be rejected because no statistically significant correlation between the use of screen reader and the time taken and number of interactions performed to activating a selection could be found.

Hypothesis 5, which proposes that the same effectiveness will be achieved across all the different hardware devices employed, can be confirmed because no statistically significant correlation between the hardware device used and the time taken and number of interactions performed to activating a selection could be found.

The number of interactions performed depends to 62 per cent on the time taken to solve the task, as shown by a strong bivariate correlation with cor = 0.78 and a very high level of statistical significance with $p = 0$. Figure 6.63 shows the number of interactions performed in correlation to the time taken as well as the linear regression of this relationship:

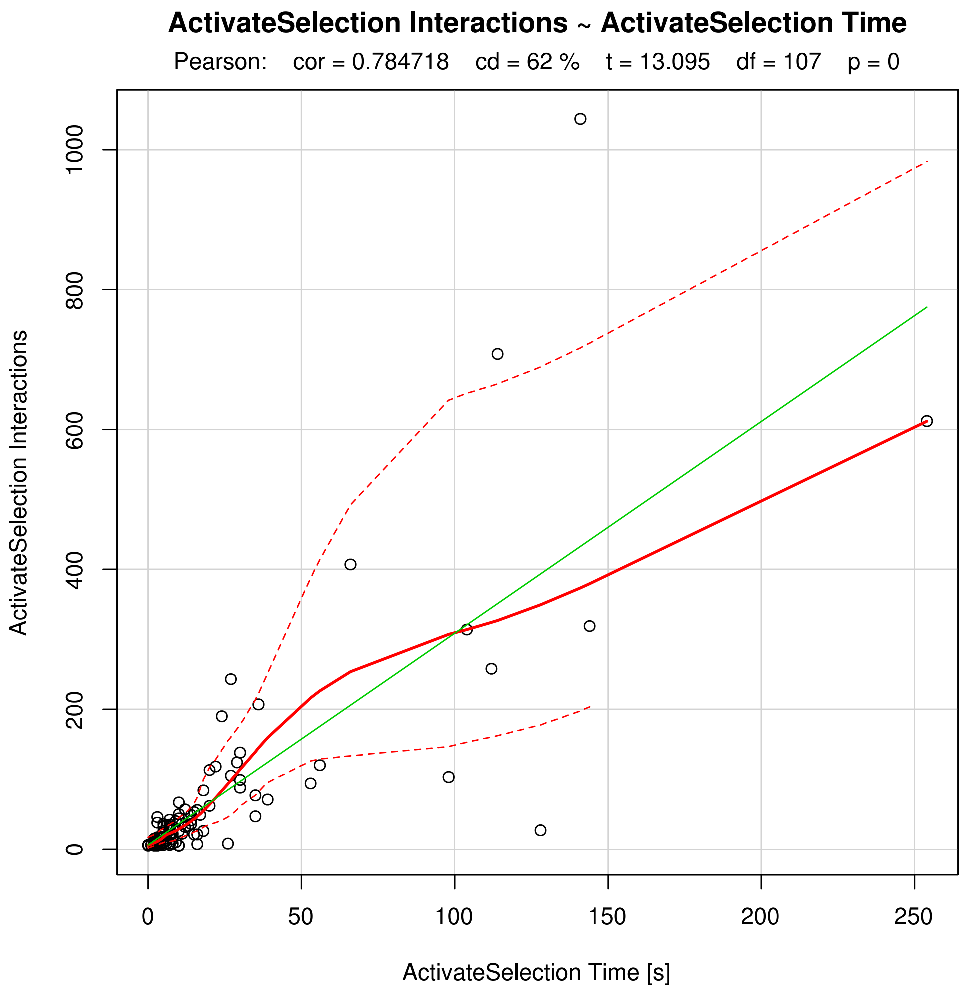


Figure 6.63: Activate Selection Interactions in correlation to Time

This correlation shows that the more time was taken, the more interactions had been performed. These results confirm Hypothesis 6, which proposes that the more effective the user interaction design concept for activating a selection is, the less time will be taken and the less interactions will be performed.

6.5.2 Deactivate Selection

In this exercise, the participants had to deactivate an arbitrary selection of the document for manipulation in order to examine if and how well the respondents are able to deactivate a selection of the document for manipulation. When a selection of the document is activated for manipulation, it can be deactivated for manipulation by unsetting the active modifier by using one of the different input modalities.

The research subjects used different input modalities for deactivating a selection. Figure 6.64 shows the different input modalities used by the persons concerned as well as the frequencies and percentages of respondents using each input modality:

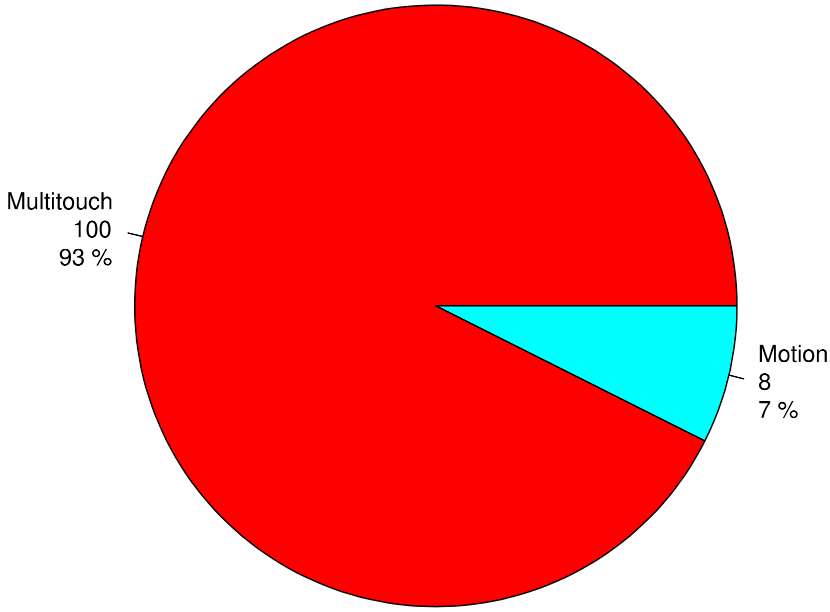


Figure 6.64: Input Modality used to Deactivate a Selection

Most of the research subjects, namely 93 per cent of them, used multitouch and only 7 per cent of the persons concerned used motion for deactivating a selection. Hypothesis 1, which proposes that blind and visually impaired persons who are using mobile devices like smart-phones and smart-tablets will choose to use the multitouch input modality for deactivating a selection, can be confirmed.

The time taken and number of interactions performed by the research subjects for deactivating a selection greatly differs between the cases. Figure 6.65 shows the central tendency, the median and dispersion as well as any outliers covering the entire time and number of interactions range as well as frequencies and percentages of the detailed distribution of the interesting ranges only without the outliers:

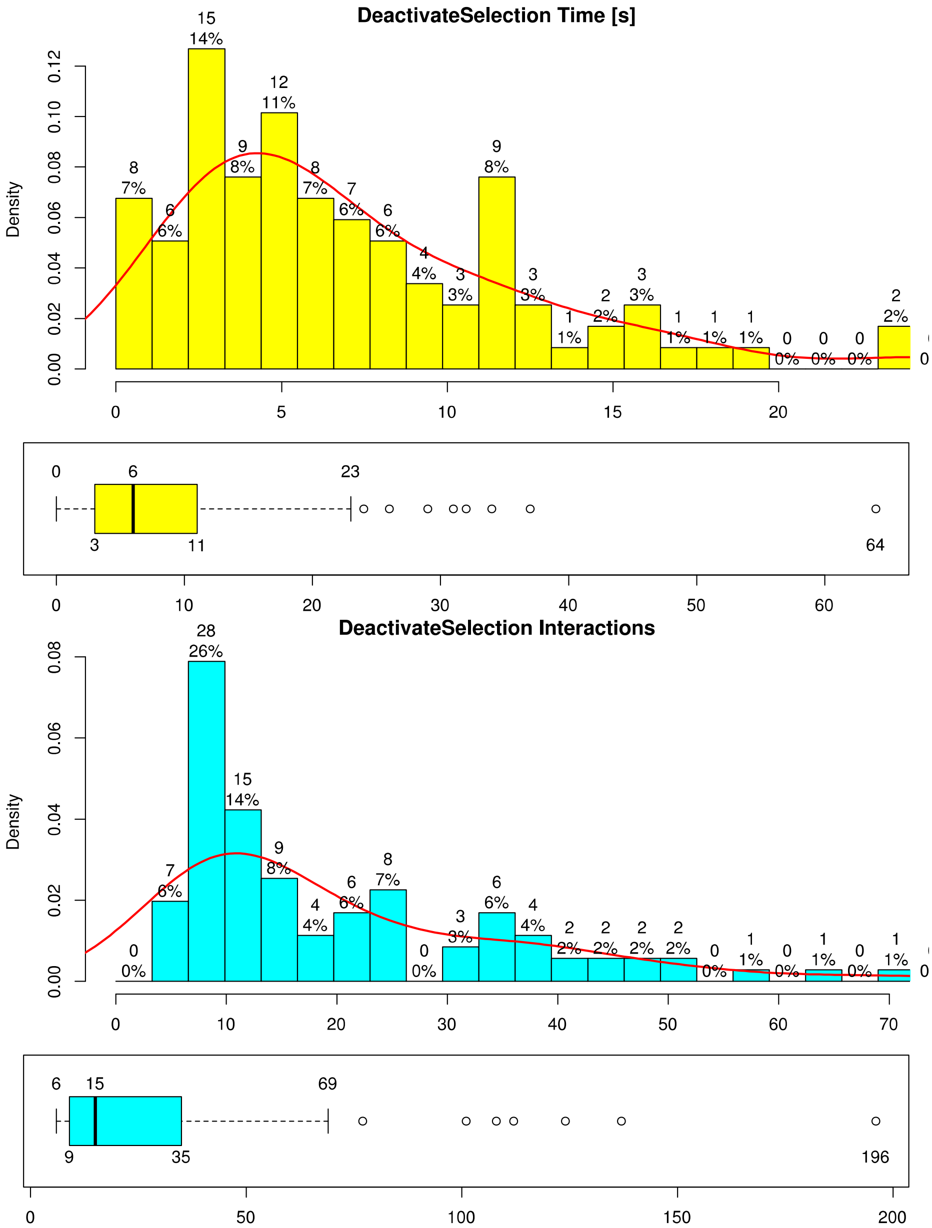


Figure 6.65: Time taken and Interactions performed to Deactivate a Selection

The middle 50 per cent of research subjects solved the task within 3 to 11 seconds and the median in 6 seconds. The fastest respondents solved the exercise in less than 1 second and the participants slower than any other, other than the outliers who might not participated seriously and took up to 64 seconds, in 23 seconds. The greatest band of persons concerned within the detailed distribution, namely 14 per cent of them, solved the task within 3 seconds.

The middle 50 per cent of research subjects solved the task by performing 9 to 35 interactions and the median by performing 15 interactions. The most effective respondents solved the exercise by performing only 6 interactions and the participants less effective than any other, other than the outliers who might not participated seriously and performed up to 196 interactions, with 69 interactions. The greatest band of persons concerned within the detailed distribution, namely 26 per cent of them, solved the task by performing between 6 and 9 interactions.

This confirms Hypothesis 2, which proposes that the user interaction design concept for deactivating a selection will be effective and enable blind and visually impaired persons to deactivate a selection of the document for manipulation within a usable amount of time taken and with a usable effort of performed interactions.

The effectiveness is dependent on the input modality used by a specific person. The time taken depends to 4 per cent on the input modality, as shown by a weak bivariate correlation with eta = 0.20 and a high level of statistical significance with p = 0.041. The number of interactions performed depends to 9 per cent on the input modality, as shown by a weak bivariate correlation with eta = 0.29 and a high level of statistical significance with p = 0.002. Figure 6.66 shows the central tendency, the median and dispersion as well as any outliers covering the entire time and number of interactions range in correlation to the different input modalities used:

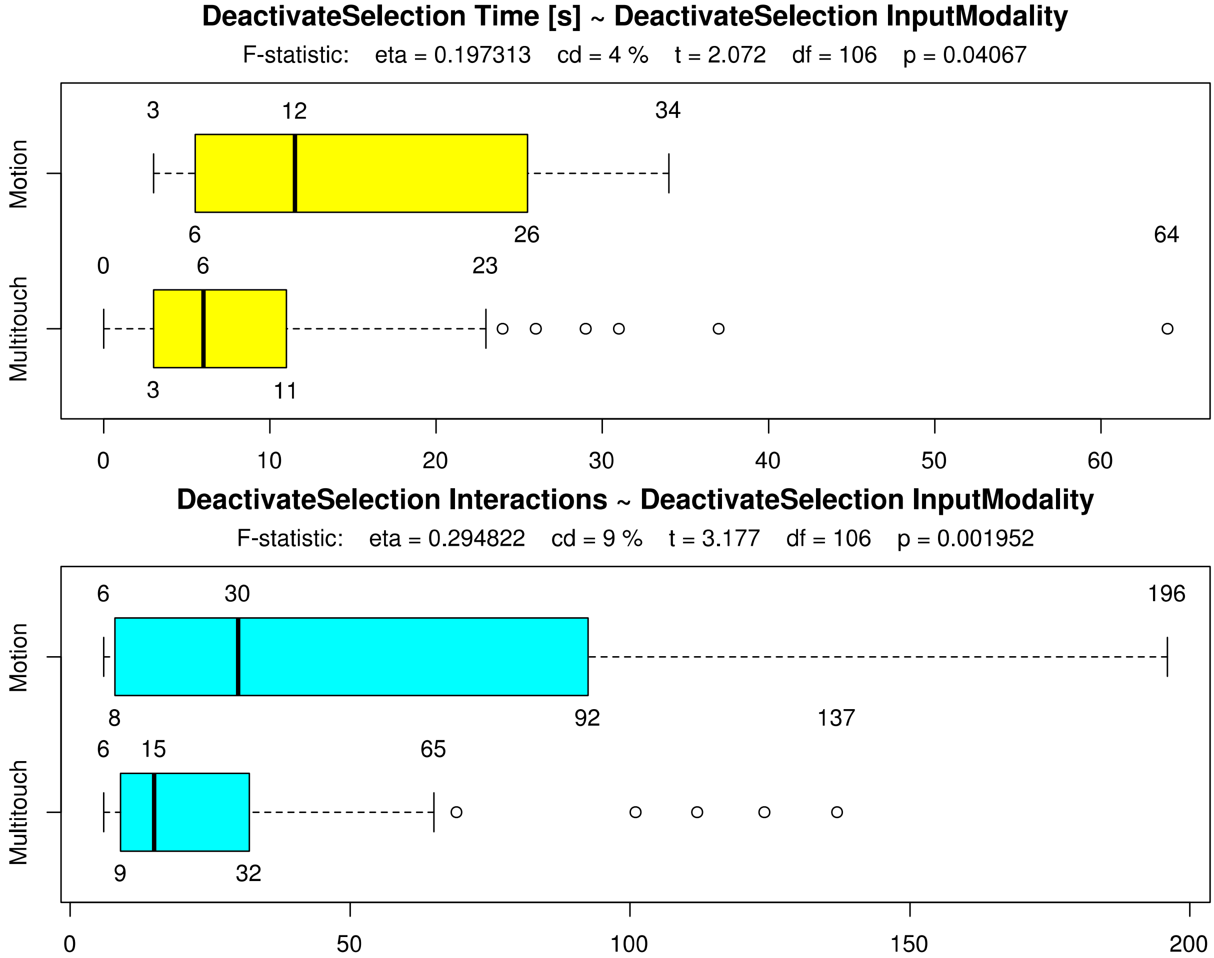


Figure 6.66: Deactivate Selection Time and Interactions in correlation to Input Modality

The middle 50 per cent of research subjects who used multitouch solved the task within 3 to 11 seconds and the median in 6 seconds. The fastest respondents solved the exercise in less than 1 second and the participants slower than any other, other than the outliers who took up to 64 seconds, in 23 seconds. On the other hand, the middle 50 per cent of research subjects who used motion solved the task within 6 to 26 seconds and the median in 12 seconds. The fastest respondents solved the exercise in 3 seconds and the participants slower than any other, other than the outliers, in 34 seconds.

The middle 50 per cent of research subjects who used multitouch solved the task by performing 9 to 32 interactions and the median by performing 15 interactions. The most effective respondents solved the exercise by performing only 6 interactions and the participants less effective than any other, other than the outliers who performed up to 137 interactions, with 65 interactions. On the other hand, the middle 50 per cent of research subjects who used motion solved the task by performing 8 to 92 interactions and the median by performing 30 interactions. The most effective respondents solved the exercise by performing only 6 interactions and the participants less effective than any other, other than the outliers, with 196 interactions.

Therefore Hypothesis 3, which proposes that the same effectiveness (time taken and number of interactions performed) to deactivating a selection will be achieved across all different input modalities, can be rejected.

Hypothesis 4, which proposes that blind and visually impaired persons who are using screen reader will perform better than respondents who are not using screen reader, can be rejected because no statistically significant correlation between the use of screen reader and the time taken and number of interactions performed to deactivating a selection could be found.

Hypothesis 5, which proposes that the same effectiveness will be achieved across all the different hardware devices employed, can be confirmed because no statistically significant correlation between the hardware device used and the time taken and number of interactions performed to deactivating a selection could be found.

The number of interactions performed depends to 57 per cent on the time taken to solve the task, as shown by a strong bivariate correlation with cor = 0.76 and a very high level of statistical significance with p = 0. Figure 6.67 shows the number of interactions performed in correlation to the time taken as well as the linear regression of this relationship:

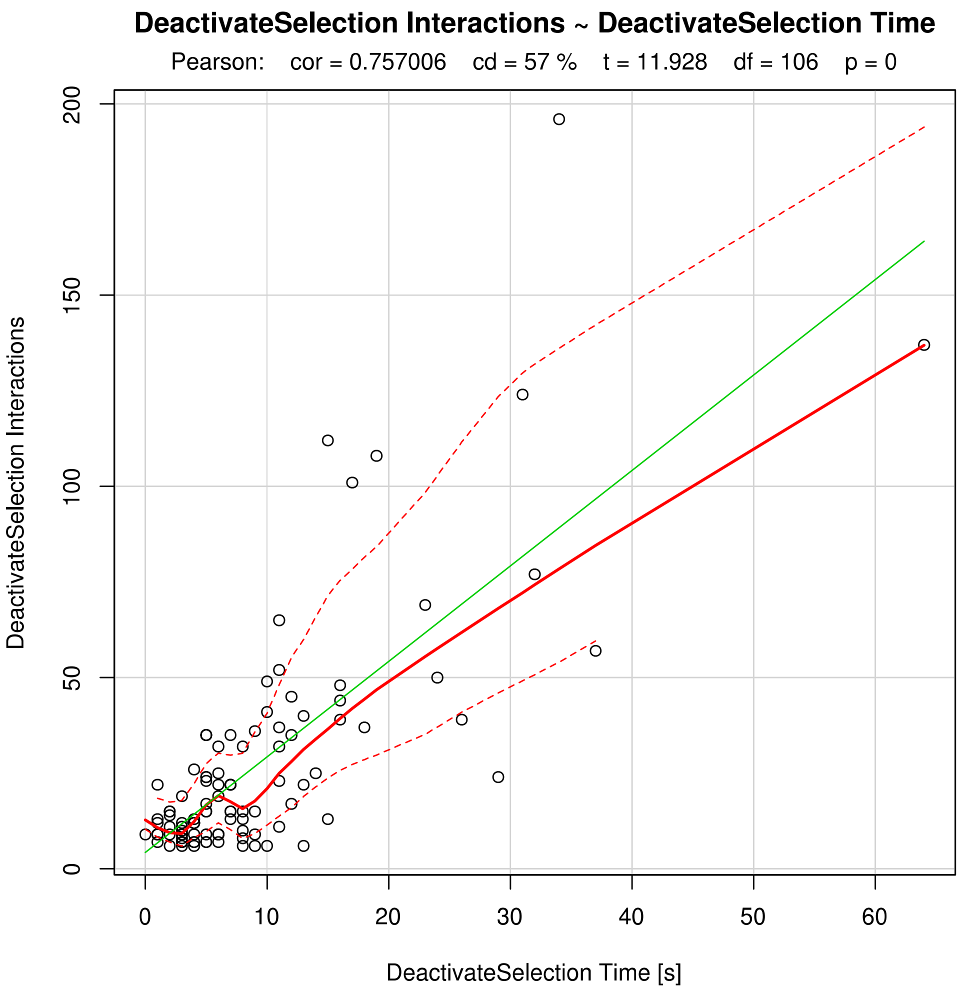


Figure 6.67: Deactivate Selection Interactions in correlation to Time

This correlation shows that the more time was taken, the more interactions had been performed. These results confirm Hypothesis 6, which proposes that the more effective the user interaction design concept for deactivating a selection is, the less time will be taken and the less interactions will be performed.

6.5.3 Move Selection

In this exercise, the participants had to move an arbitrary selection of the document to another position within the document in order to examine if and how well the respondents are able to move a selection of the document to another position within the document. When a selection of the document is activated for manipulation, it can be moved by focusing the target position within the document to where the selection should be moved and unsetting the active modifier by using one of the different input modalities.

The research subjects used different input modalities for moving a selection. Figure 6.68 shows the different input modalities used by the persons concerned as well as the frequencies and percentages of respondents using each input modality:

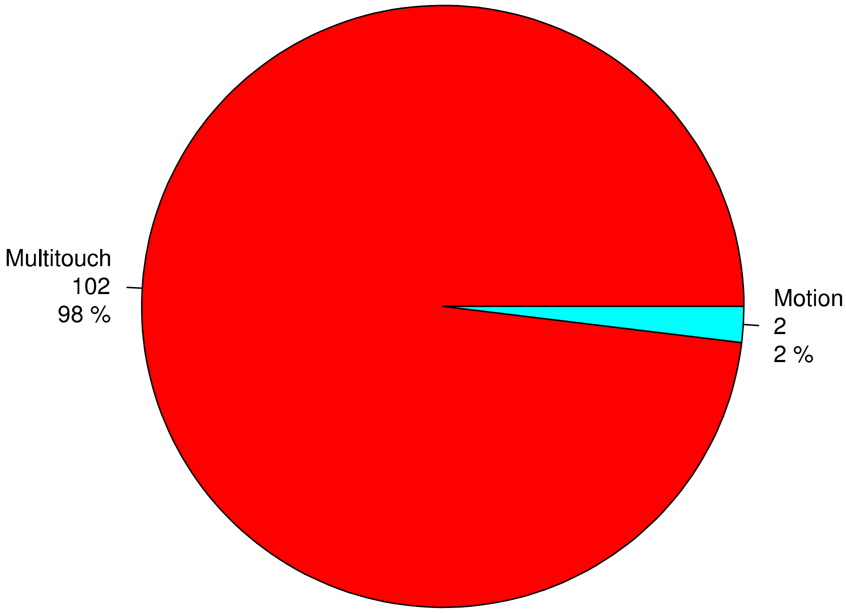


Figure 6.68: Input Modality used to Move a Selection

Most of the research subjects, namely 98 per cent of them, used multitouch and only 2 per cent of the persons concerned used motion for moving a selection. Hypothesis 1, which proposes that blind and visually impaired persons who are using mobile devices like smart-phones and smart-tablets will choose to use the multitouch input modality for moving a selection, can be confirmed.

The time taken and number of interactions performed by the research subjects for moving a selection greatly differs between the cases. Figure 6.69 shows the central tendency, the median and dispersion as well as any outliers covering the entire time and number of interactions range as well as frequencies and percentages of the detailed distribution of the interesting ranges only without the outliers:

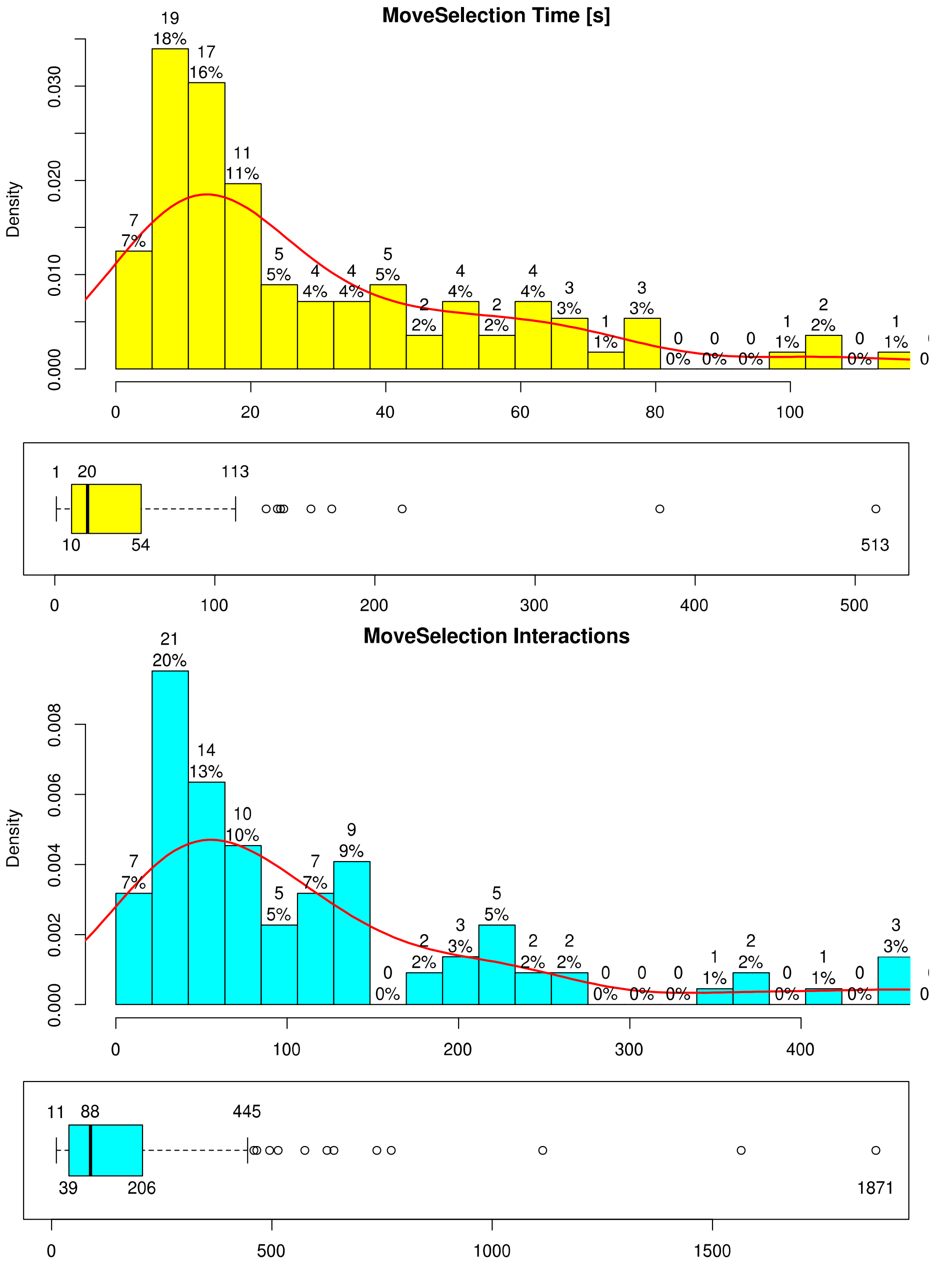


Figure 6.69: Time taken and Interactions performed to Move a Selection

The middle 50 per cent of research subjects solved the task within 10 to 54 seconds and the median in 20 seconds. The fastest respondents solved the exercise in 1 second and the participants slower than any other, other than the outliers who might not participated seriously and took up to 513 seconds, in 113 seconds. The greatest band of persons concerned within the detailed distribution, namely 18 per cent of them, solved the task within 5 and 10 seconds.

The middle 50 per cent of research subjects solved the task by performing 39 to 206 interactions and the median by performing 88 interactions. The most effective respondents solved the exercise by performing only 11 interactions and the participants less effective than any other, other than the outliers who might not participated seriously and performed up to 1871 interactions, with 445 interactions. The greatest band of persons concerned within the detailed distribution, namely 20 per cent of them, solved the task by performing between 10 and 20 interactions.

This confirms Hypothesis 2, which proposes that the user interaction design concept for moving a selection will be effective and enable blind and visually impaired persons to move a selection of the document to another position within the document within a usable amount of time taken and with a usable effort of performed interactions.

Hypothesis 3, which proposes that the same effectiveness (time taken and number of interactions performed) will be achieved across all different input modalities, can be confirmed because no statistically significant correlation between the input modality used and the time taken and number of interactions performed to moving a selection could be found.

Hypothesis 4, which proposes that blind and visually impaired persons who are using screen reader will perform better than respondents who are not using screen reader, can be rejected because no statistically significant correlation between the use of screen reader and the time taken and number of interactions performed to moving a selection could be found.

Hypothesis 5, which proposes that the same effectiveness will be achieved across all the different hardware devices employed, can be confirmed because no statistically significant correlation between the hardware device used and the time taken and number of interactions performed to moving a selection could be found.

The number of interactions performed depends to 72 per cent on the time taken to solve the task, as shown by a very strong bivariate correlation with cor = 0.85 and a very high level of statistical significance with p = 0. Figure 6.70 shows the number of interactions performed in correlation to the time taken as well as the linear regression of this relationship:

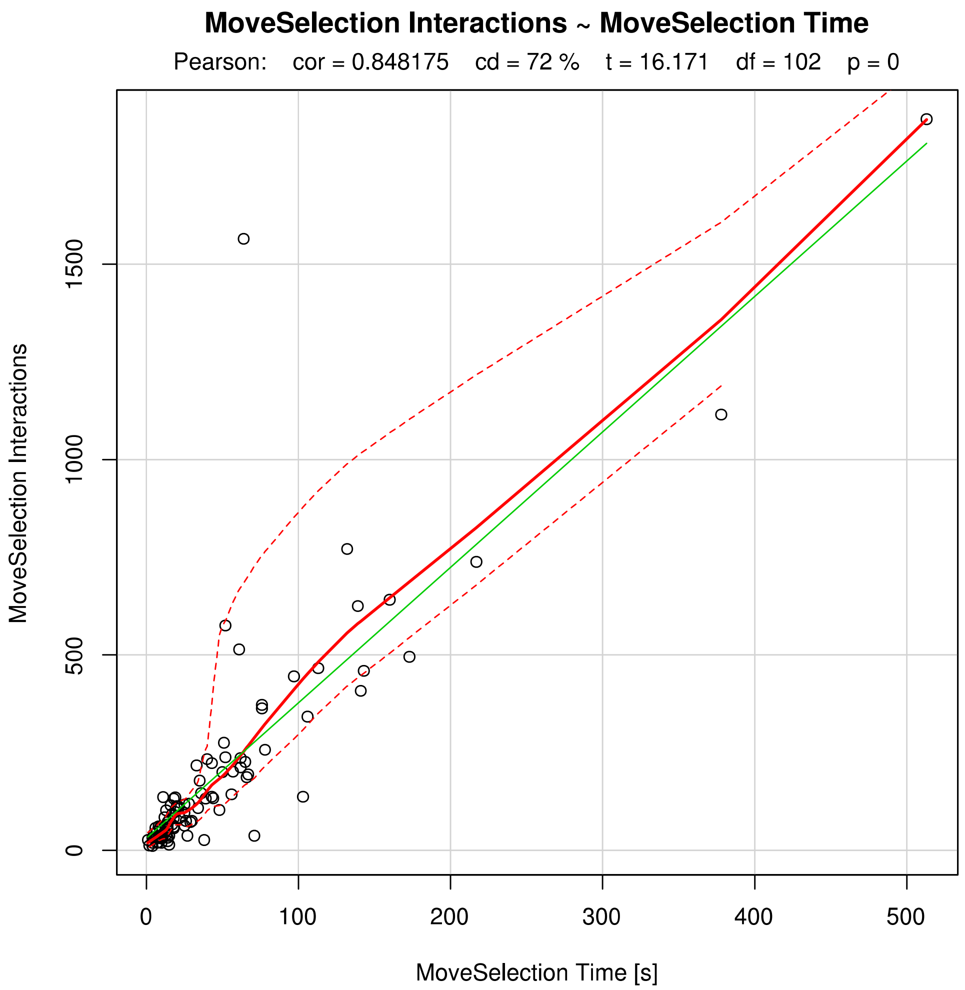


Figure 6.70: Move Selection Interactions in correlation to Time

This correlation shows that the more time was taken, the more interactions had been performed. These results confirm Hypothesis 6, which proposes that the more effective the user interaction design concept for moving a selection is, the less time will be taken and the less interactions will be performed.

6.5.4 Remove Selection

In this exercise, the participants had to remove an arbitrary selection of the document from the document in order to examine if and how well the respondents are able to remove a selection of the document from the document. When a selection of the document is activated for manipulation, it can be removed by unsetting the focus from the structure element which currently has the focus and unsetting the active modifier by using one of the different output modalities.

The research subjects used different input modalities for removing a selection. Figure 6.71 shows the different input modalities used by the persons concerned as well as the frequencies and percentages of respondents using each input modality:

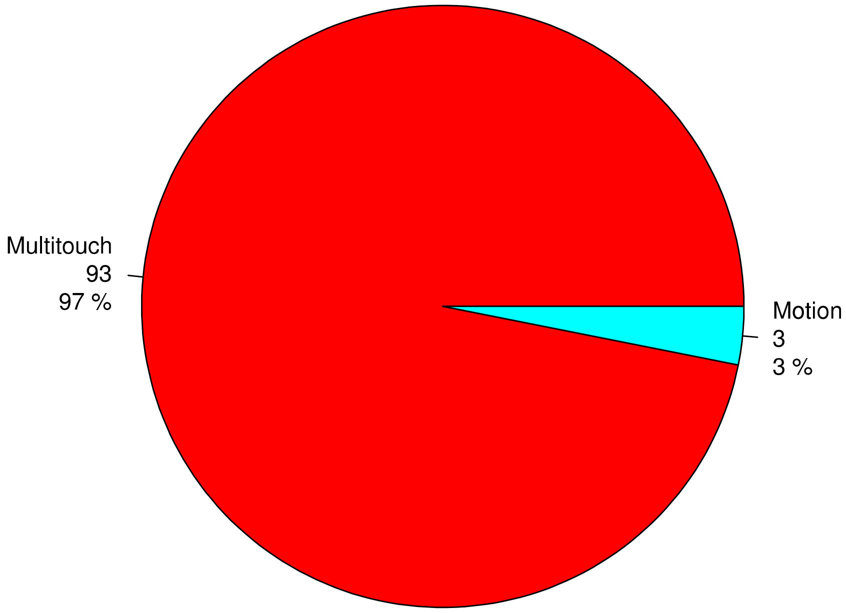


Figure 6.71: Input Modality used to Remove a Selection

Most of the research subjects, namely 97 per cent of them, used multitouch and only 3 per cent of the persons concerned used motion for removing a selection. Hypothesis 1, which proposes that blind and visually impaired persons who are using mobile devices like smart-phones and smart-tablets will choose to use the multitouch input modality for removing a selection, can be confirmed.

The time taken and number of interactions performed by the research subjects for removing a selection greatly differs between the cases. Figure 6.72 shows the central tendency, the median and dispersion as well as any outliers covering the entire time and number of interactions range as well as frequencies and percentages of the detailed distribution of the interesting ranges only without the outliers:

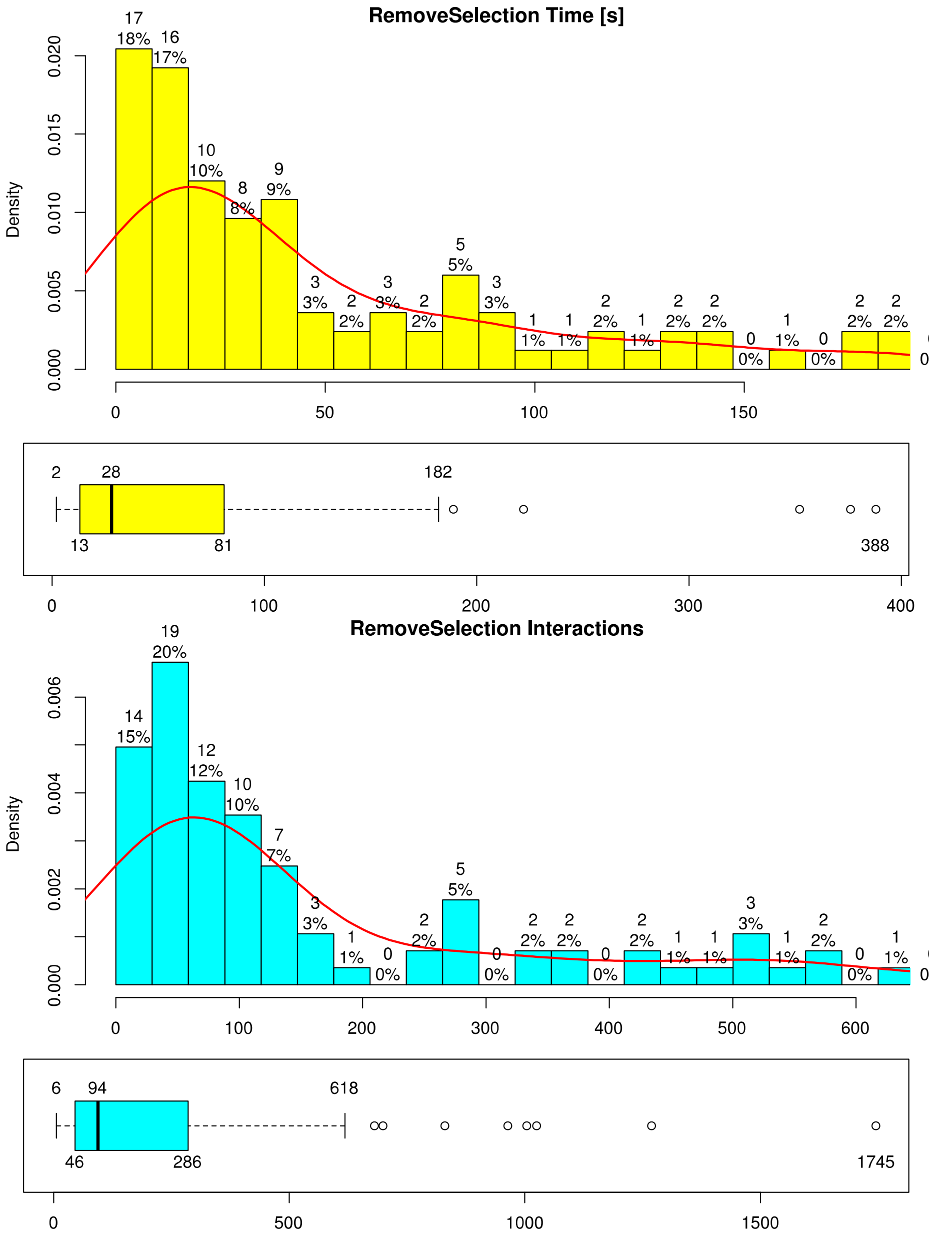


Figure 6.72: Time taken and Interactions performed to Remove a Selection

The middle 50 per cent of research subjects solved the task within 13 to 81 seconds and the median in 28 seconds. The fastest respondents solved the exercise in 2 seconds and the participants slower than any other, other than the outliers who might not participated seriously and took up to 388 seconds, in 182 seconds. The greatest band of persons concerned within the detailed distribution, namely 18 per cent of them, solved the task within 0 and 9 seconds.

The middle 50 per cent of research subjects solved the task by performing 46 to 286 interactions and the median by performing 94 interactions. The most effective respondents solved the exercise by performing only 6 interactions and the participants less effective than any other, other than the outliers who might not participated seriously and performed up to 1745 interactions, with 618 interactions. The greatest band of persons concerned within the detailed distribution, namely 20 per cent of them, solved the task by performing between 30 and 60 interactions.

This confirms Hypothesis 2, which proposes that the user interaction design concept for removing a selection will be effective and enable blind and visually impaired persons to remove a selection of the document from the document within a usable amount of time taken and with a usable effort of performed interactions.

The effectiveness is dependent on the input modality used by a specific person. The time taken depends to 5 per cent on the input modality, as shown by a weak bivariate correlation with eta = 0.23 and a high level of statistical significance with p = 0.025. The number of interactions performed depends to 7 per cent on the input modality, as shown by a weak bivariate correlation with eta = 0.26 and a high level of statistical significance with p = 0.0099. Figure 6.73 shows the central tendency, the median and dispersion as well as any outliers covering the entire time and number of interactions range in correlation to the different input modalities used:

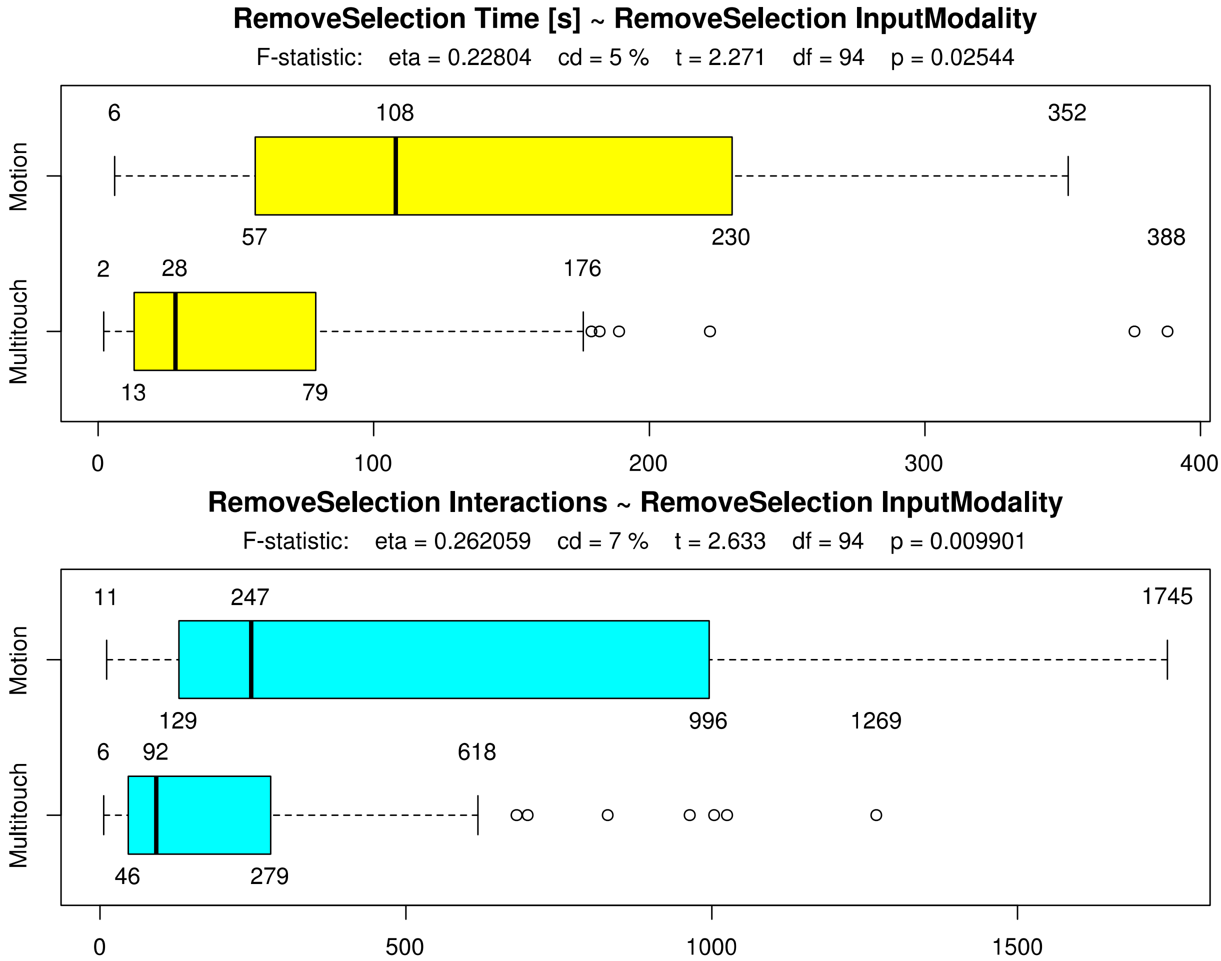


Figure 6.73: Remove Selection Time and Interactions in correlation to Input Modality

The middle 50 per cent of research subjects who used multitouch solved the task within 13 to 79 seconds and the median in 28 seconds. The fastest respondents solved the exercise in 2 seconds and the participants slower than any other, other than the outliers who took up to 388 seconds, in 176 seconds. On the other hand, the middle 50 per cent of research subjects who used motion solved the task within 57 to 230 seconds and the median in 108 seconds. The fastest respondents solved the exercise in 6 seconds and the participants slower than any other, other than the outliers, in 352 seconds.

The middle 50 per cent of research subjects who used multitouch solved the task by performing 46 to 279 interactions and the median by performing 92 interactions. The most effective respondents solved the exercise by performing only 6 interactions and the participants less effective than any other, other than the outliers who performed up to 1269 interactions, with 618 interactions. On the other hand, the middle 50 per cent of research subjects who used motion solved the task by performing 129 to 996 interactions and the median by performing 247 interactions. The most effective respondents solved the exercise by performing only 11 interactions and the participants less effective than any other, other than the outliers, with 1745 interactions.

Therefore Hypothesis 3, which proposes that the same effectiveness (time taken and number of interactions performed) to removing a selection will be achieved across all different input modalities, can be rejected.

Hypothesis 4, which proposes that blind and visually impaired persons who are using screen reader will perform better than respondents who are not using screen reader, can be rejected because no statistically significant correlation between the use of screen reader and the time taken and number of interactions performed to removing a selection could be found.

Hypothesis 5, which proposes that the same effectiveness will be achieved across all the different hardware devices employed, can be confirmed because no statistically significant correlation between the hardware device used and the time taken and number of interactions performed to removing a selection could be found.

The number of interactions performed depends to 76 per cent on the time taken to solve the task, as shown by a very strong bivariate correlation with cor = 0.87 and a very high level of statistical significance with p = 0. Figure 6.74 shows the number of interactions performed in correlation to the time taken as well as the linear regression of this relationship:

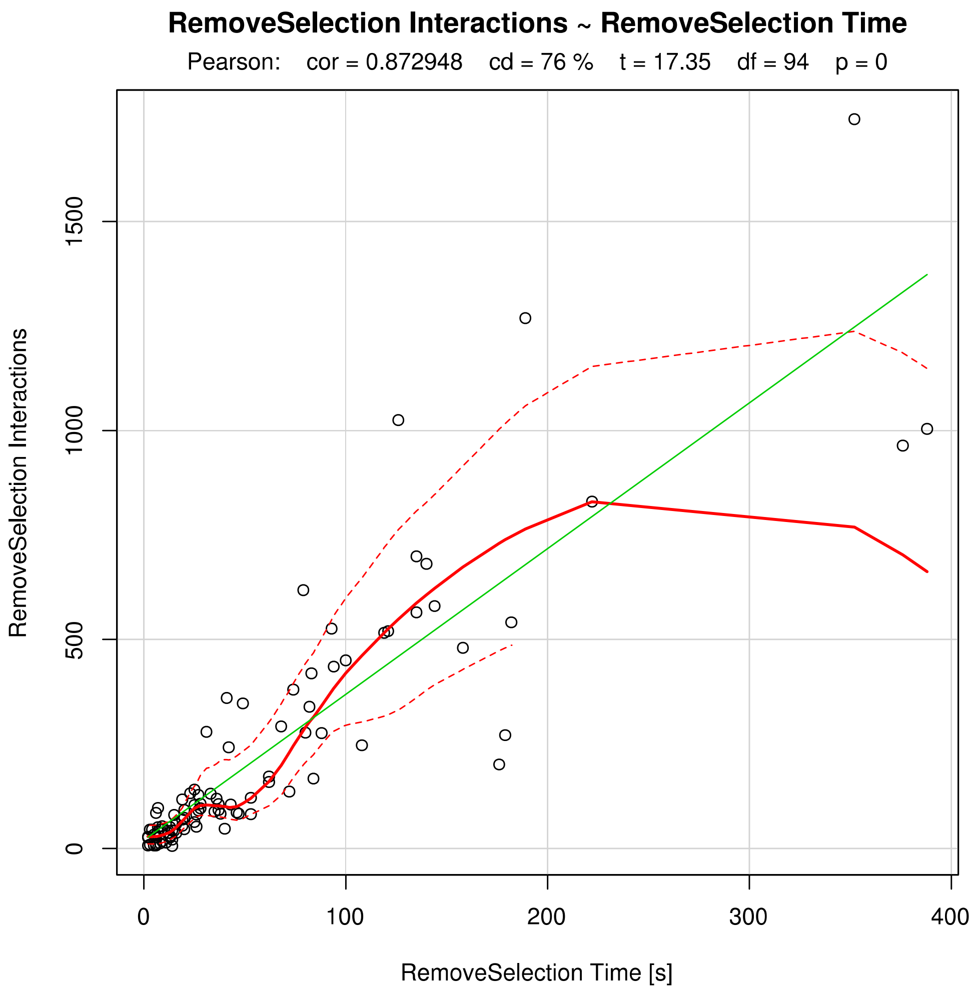


Figure 6.74: Remove Selection Interactions in correlation to Time

This correlation shows that the more time was taken, the more interactions had been performed. These results confirm Hypothesis 6, which proposes that the more effective the user interaction design concept for removing a selection is, the less time will be taken and the less interactions will be performed.

6.5.5 Insert Element

In this exercise, the participants had to insert a new structure element into the document at an arbitrary position in order to examine if and how well the respondents are able to insert a new structure element into the document. When the position within the document after which the new structure element should be inserted is focused, it can be inserted by setting the insertElement cursor to a position on the insertElement axis by using one of the different input modalities, optionally moving the insertElement cursor on the insertElement axis forward or backward to choose another element type and setting the select modifier by using one of the different input modalities.

The research subjects used different input modalities for inserting a new structure element. Figure 6.75 shows the different input modalities used by the persons concerned as well as the frequencies and percentages of respondents using each input modality:

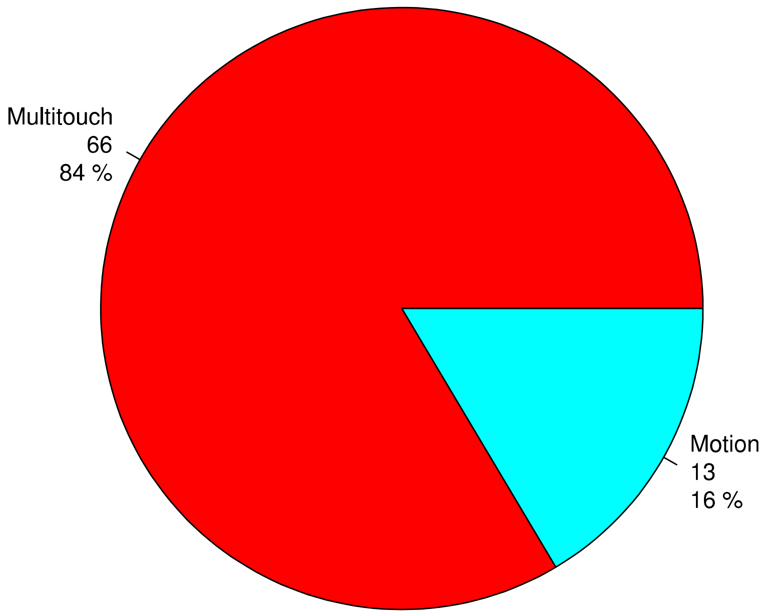


Figure 6.75: Input Modality used to Insert an Element

Most of the research subjects, namely 84 per cent of them, used multitouch and only 16 per cent of the persons concerned used motion for inserting a new structure element. Hypothesis 1, which proposes that blind and visually impaired persons who are using mobile devices like smart-phones and smart-tablets will choose to use the multitouch input modality for inserting a new structure element, can be confirmed.

The time taken and number of interactions performed by the research subjects for inserting a new structure element greatly differs between the cases. Figure 6.76 shows the central tendency, the median and dispersion as well as any outliers covering the entire time and number of interactions range as well as frequencies and percentages of the detailed distribution of the interesting ranges only without the outliers:

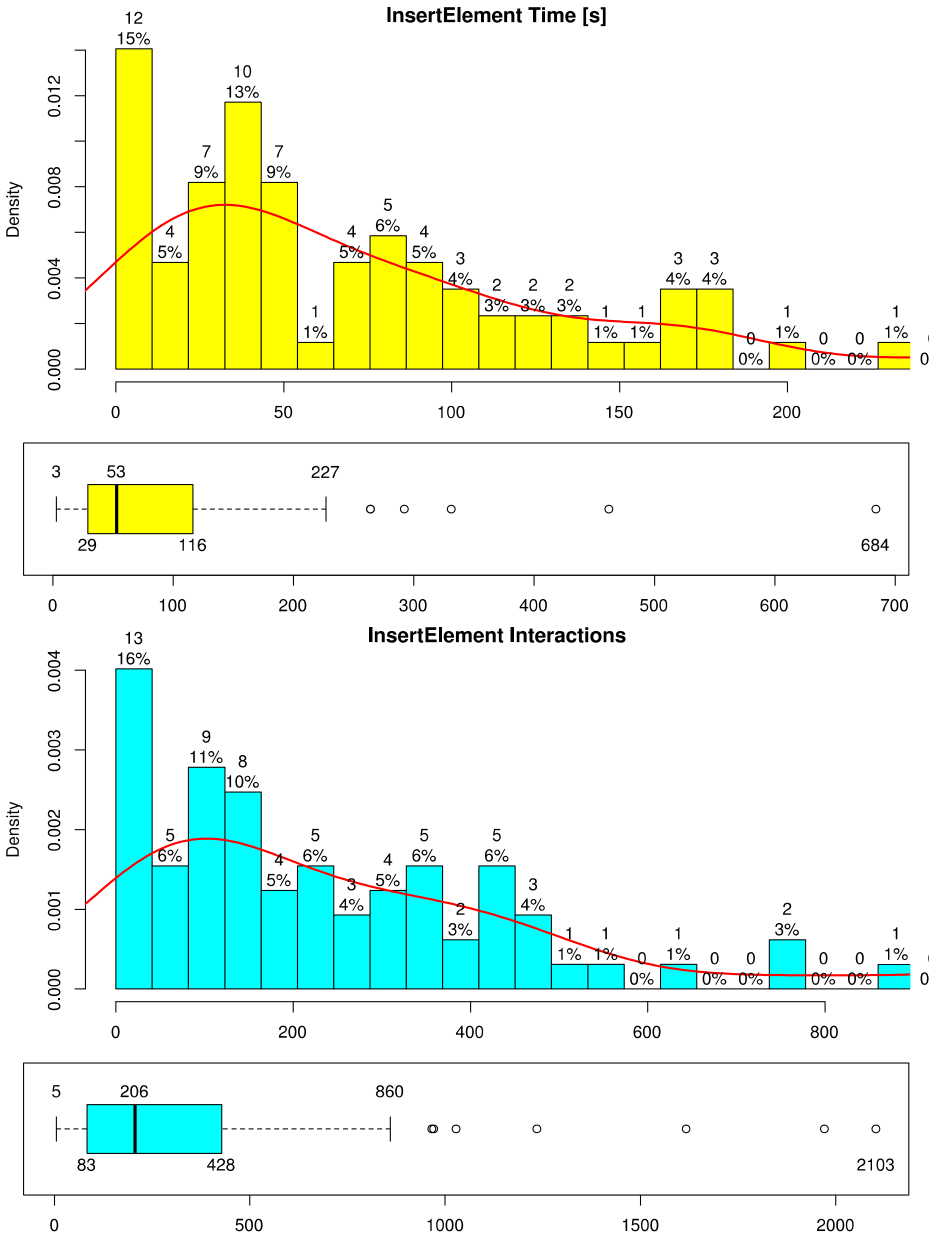


Figure 6.76: Time taken and Interactions performed to Insert an Element

The middle 50 per cent of research subjects solved the task within 29 to 116 seconds and the median in 53 seconds. The fastest respondents solved the exercise in 3 seconds and the participants slower than any other, other than the outliers who might not participated seriously and took up to 684 seconds, in 227 seconds. The greatest band of persons concerned within the detailed distribution, namely 15 per cent of them, solved the task within 0 and 10 seconds.

The middle 50 per cent of research subjects solved the task by performing 83 to 428 interactions and the median by performing 206 interactions. The most effective respondents solved the exercise by performing only 5 interactions and the participants less effective than any other, other than the outliers who might not participated seriously and performed up to 2103 interactions, with 860 interactions. The greatest band of persons concerned within the detailed distribution, namely 16 per cent of them, solved the task by performing between 0 and 40 interactions.

This confirms Hypothesis 2, which proposes that the user interaction design concept for inserting a new structure element will be effective and enable blind and visually impaired persons to insert a new structure element into the document within a usable amount of time taken and with a usable effort of performed interactions.

The effectiveness is dependent on the input modality used by a specific person. The time taken depends to 5 per cent on the input modality, as shown by a weak bivariate correlation with eta = 0.22 and a high level of statistical significance with p = 0.054. The number of interactions performed depends to 6 per cent on the input modality, as shown by a weak bivariate correlation with eta = 0.25 and a high level of statistical significance with p = 0.027. Figure 6.77 shows the central tendency, the median and dispersion as well as any outliers covering the entire time and number of interactions range in correlation to the different input modalities used:

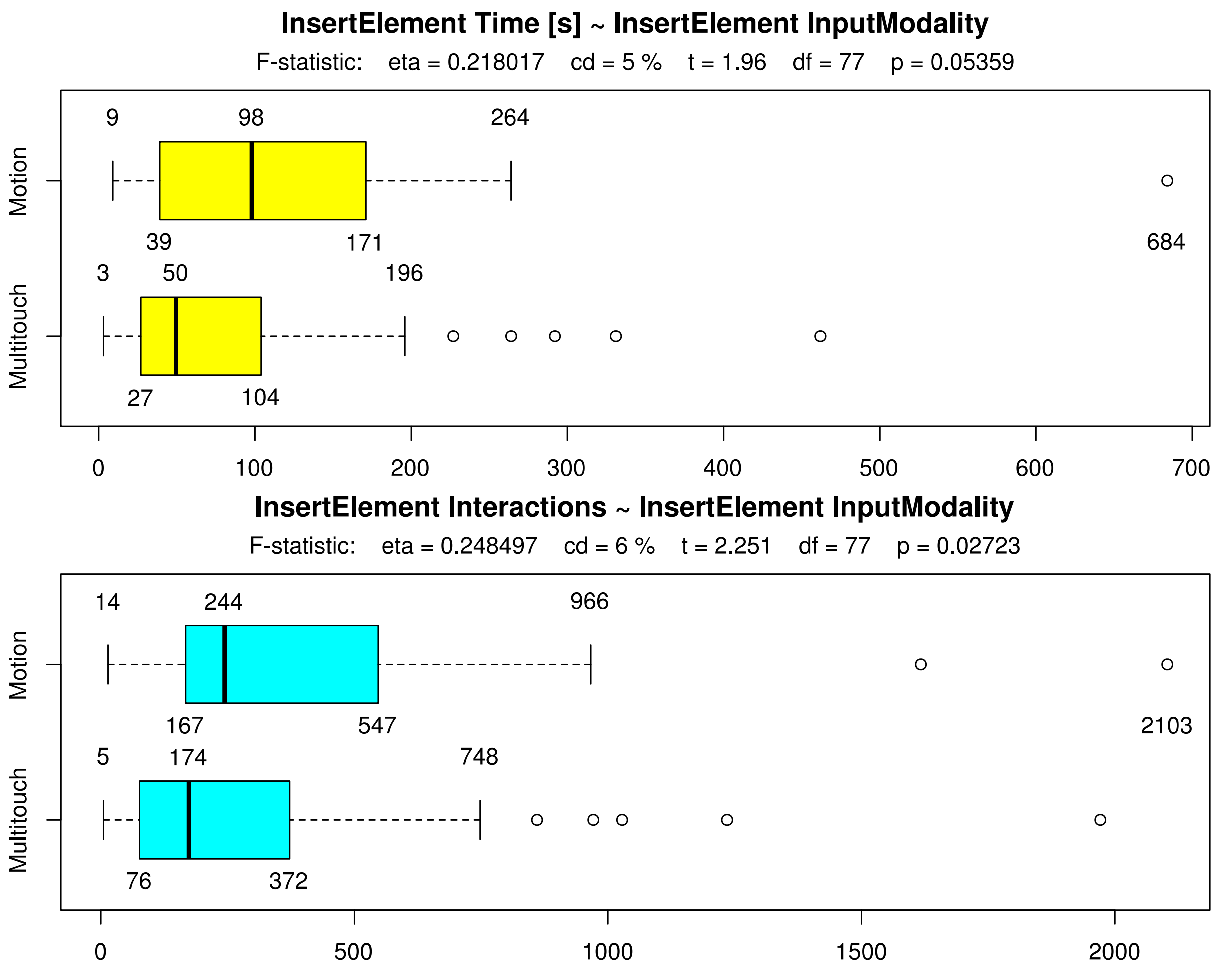


Figure 6.77: Insert Element Time and Interactions in correlation to Input Modality

The middle 50 per cent of research subjects who used multitouch solved the task within 27 to 104 seconds and the median in 50 seconds. The fastest respondents solved the exercise in 3 seconds and the participants slower than any other, other than the outliers, in 196 seconds. On the other hand, the middle 50 per cent of research subjects who used motion solved the task within 39 to 171 seconds and the median in 98 seconds. The fastest respondents solved the exercise in 9 seconds and the participants slower than any other, other than the outliers who took up to 684 seconds, in 264 seconds.

The middle 50 per cent of research subjects who used multitouch solved the task by performing 76 to 372 interactions and the median by performing 174 interactions. The most effective respondents solved the exercise by performing only 5 interactions and the participants less effective than any other, other than the outliers, with 748 interactions. On the other hand, the middle 50 per cent of research subjects who used motion solved

the task by performing 167 to 547 interactions and the median by performing 244 interactions. The most effective respondents solved the exercise by performing only 14 interactions and the participants less effective than any other, other than the outliers who performed up to 2103 interactions, with 966 interactions.

Therefore Hypothesis 3, which proposes that the same effectiveness (time taken and number of interactions performed) to inserting a new structure element will be achieved across all different input modalities, can be rejected.

Hypothesis 4, which proposes that blind and visually impaired persons who are using screen reader will perform better than respondents who are not using screen reader, can be rejected because no statistically significant correlation between the use of screen reader and the time taken and number of interactions performed to inserting a new structure element could be found.

Hypothesis 5, which proposes that the same effectiveness will be achieved across all the different hardware devices employed, can be confirmed because no statistically significant correlation between the hardware device used and the time taken and number of interactions performed to inserting a new structure element could be found.

The number of interactions performed depends to 71 per cent on the time taken to solve the task, as shown by a very strong bivariate correlation with cor = 0.84 and a very high level of statistical significance with p = 0. Figure 6.78 shows the number of interactions performed in correlation to the time taken as well as the linear regression of this relationship:

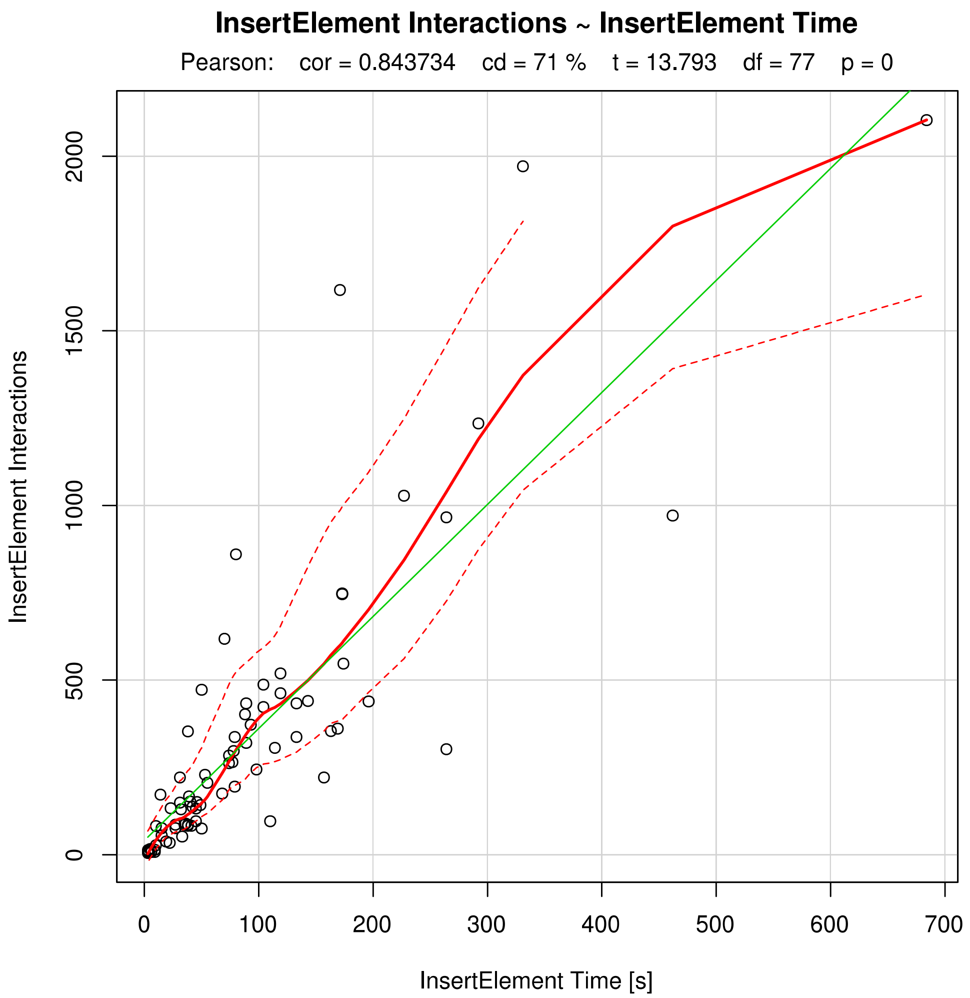


Figure 6.78: Insert Element Interactions in correlation to Time

This correlation shows that the more time was taken, the more interactions had been performed. These results confirm Hypothesis 6, which proposes that the more effective the user interaction design concept for inserting a new structure element is, the less time will be taken and the less interactions will be performed.

6.5.6 Insert Text

In this exercise, the participants had to insert a new text into the document at an arbitrary position in order to examine if and how well the respondents are able to insert a new text into the document. When the position within the document after which the new text should be inserted is focused, it can be inserted by setting the insertText modifier by using one of the different input modalities, speaking the text to insert and unsetting the insertText modifier by using one of the different input modalities.

The research subjects used different input modalities for inserting a new text. Figure 6.79 shows the different input modalities used by the persons concerned as well as the frequencies and percentages of respondents using each input modality:

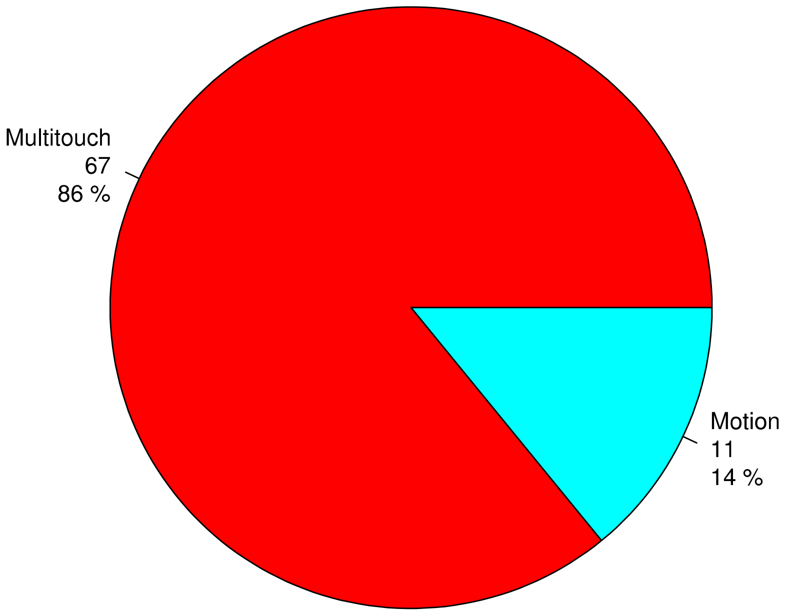


Figure 6.79: Input Modality used to Insert a Text

Most of the research subjects, namely 86 per cent of them, used multitouch and only 14 per cent of the persons concerned used motion for inserting a new text. Hypothesis 1, which proposes that blind and visually impaired persons who are using mobile devices like smart-phones and smart-tablets will choose to use the multitouch input modality for inserting a new text, can be confirmed.

The time taken and number of interactions performed by the research subjects for inserting a new text greatly differs between the cases. Figure 6.80 shows the central tendency, the median and dispersion as well as any outliers covering the entire time and number of interactions range as well as frequencies and percentages of the detailed distribution of the interesting ranges only without the outliers:

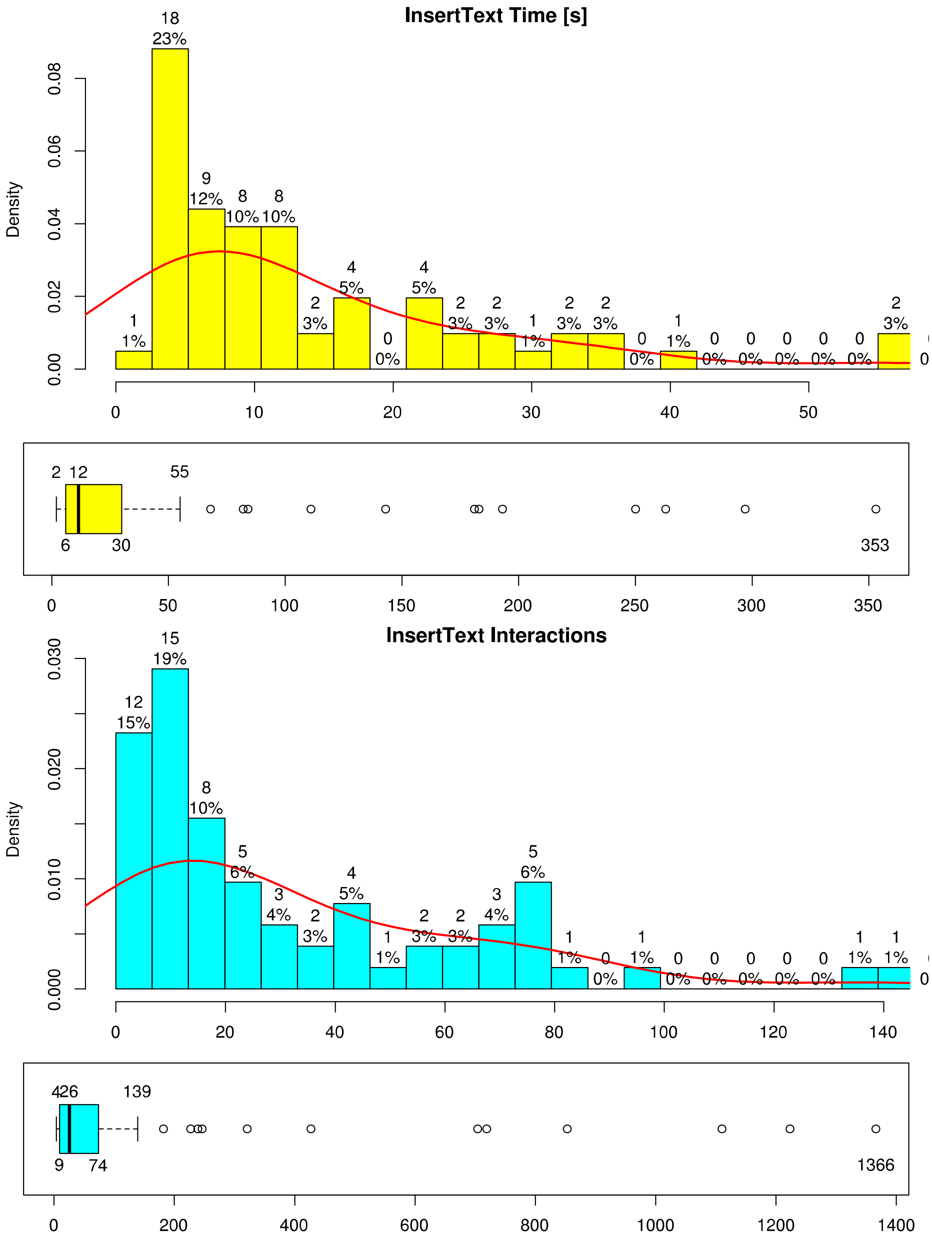


Figure 6.80: Time taken and Interactions performed to Insert a Text

The middle 50 per cent of research subjects solved the task within 6 to 30 seconds and the median in 12 seconds. The fastest respondents solved the exercise in 2 seconds and the participants slower than any other, other than the outliers who might not participated seriously and took up to 353 seconds, in 55 seconds. The greatest band of person s concerned within the detailed distribution, namely 23 per cent of them, solved the task within 2 and 5 seconds.

The middle 50 per cent of research subjects solved the task by performing 9 to 74 interactions and the median by performing 26 interactions. The most effective respondents solved the exercise by performing only 4 interactions and the participants less effective than any other, other than the outliers who might not participated seriously and performed up to 1366 interactions, with 139 interactions. The greatest band of persons concerned within the detailed distribution, namely 19 per cent of them, solved the task by performing between 7 and 14 interactions.

This confirms Hypothesis 2, which proposes that the user interaction design concept for inserting a new text will be effective and enable blind and visually impaired persons to insert a new text into the document within a usable amount of time taken and with a usable effort of performed interactions.

The effectiveness is dependent on the input modality used by a specific person. The time taken depends to 6 per cent on the input modality, as shown by a weak bivariate correlation with eta = 0.25 and a high level of statistical significance with p = 0.025. The number of interactions performed depends to 7 per cent on the input modality, as shown by a weak bivariate correlation with eta = 0.27 and a high level of statistical significance with p = 0.015. Figure 6.81 shows the central tendency, the median and dispersion as well as any outliers covering the entire time and number of interactions range in correlation to the different input modalities used:

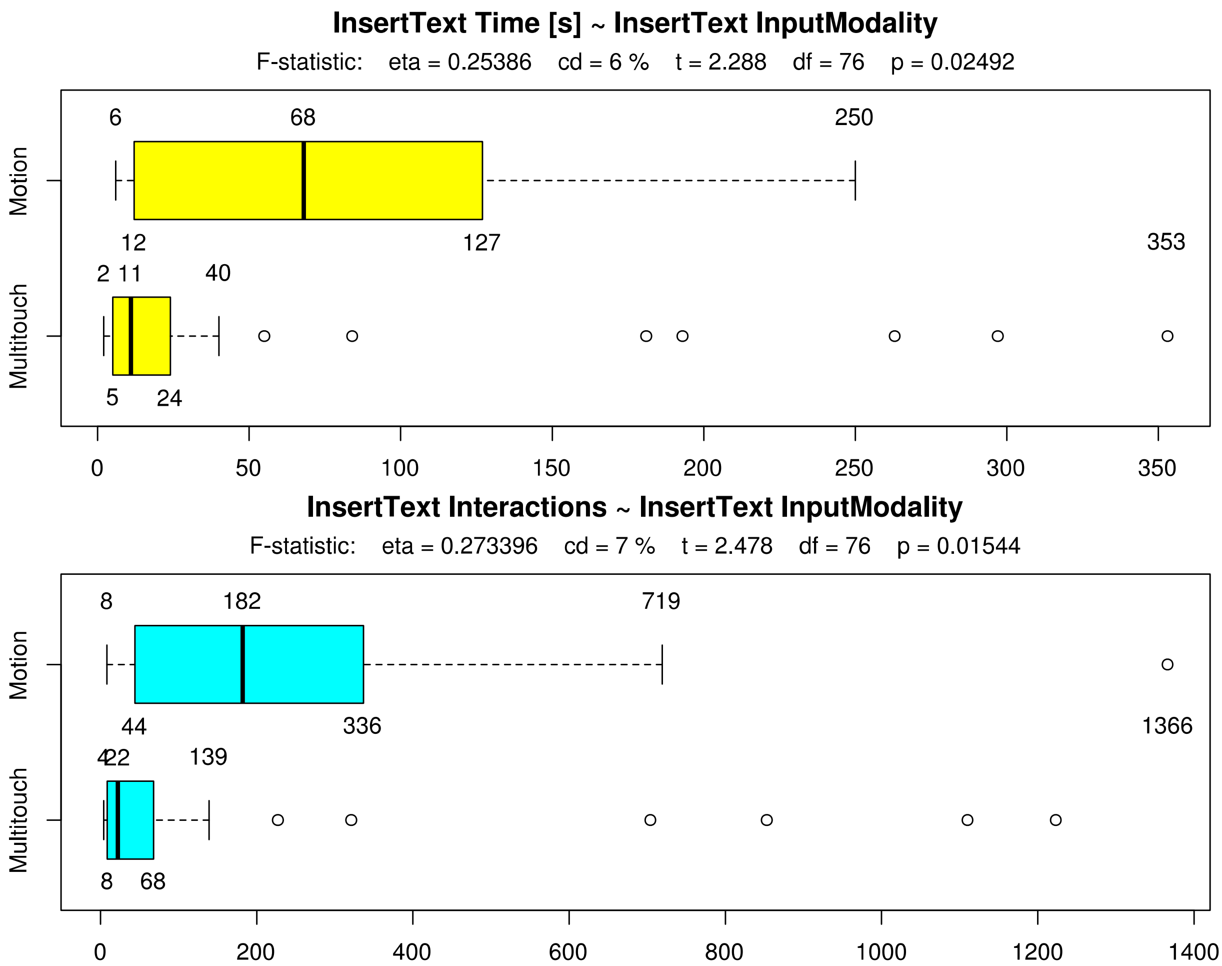


Figure 6.81: Insert Text Time and Interactions in correlation to Input Modality

The middle 50 per cent of research subjects who used multitouch solved the task within 5 to 24 seconds and the median in 11 seconds. The fastest respondents solved the exercise in 2 seconds and the participants slower than any other, other than the outliers who took up to 353 seconds, in 40 seconds. On the other hand, the middle 50 per cent of research subjects who used motion solved the task within 12 to 127 seconds and the median in 68 seconds. The fastest respondents solved the exercise in 68 seconds and the participants slower than any other, other than the outliers, in 250 seconds.

The middle 50 per cent of research subjects who used multitouch solved the task by performing 8 to 68 interactions and the median by performing 22 interactions. The most effective respondents solved the exercise by performing only 4 interactions and the participants less effective than any other, other than the outliers, with 139 interactions. On the other hand, the middle 50 per cent of research subjects who used motion solved

the task by performing 44 to 336 interactions and the median by performing 182 interactions. The most effective respondents solved the exercise by performing only 8 interactions and the participants less effective than any other, other than the outliers who performed up to 1366 interactions, with 719 interactions.

Therefore Hypothesis 3, which proposes that the same effectiveness (time taken and number of interactions performed) to inserting a new text will be achieved across all different input modalities, can be rejected.

Hypothesis 4, which proposes that blind and visually impaired persons who are using screen reader will perform better than respondents who are not using screen reader, can be rejected because no statistically significant correlation between the use of screen reader and the time taken and number of interactions performed to inserting a new text could be found.

Hypothesis 5, which proposes that the same effectiveness will be achieved across all the different hardware devices employed, can be confirmed because no statistically significant correlation between the hardware device used and the time taken and number of interactions performed to inserting a new text could be found.

The number of interactions performed depends to 71 per cent on the time taken to solve the task, as shown by a very strong bivariate correlation with cor = 0.84 and a very high level of statistical significance with $p = 0$. Figure 6.82 shows the number of interactions performed in correlation to the time taken as well as the linear regression of this relationship:

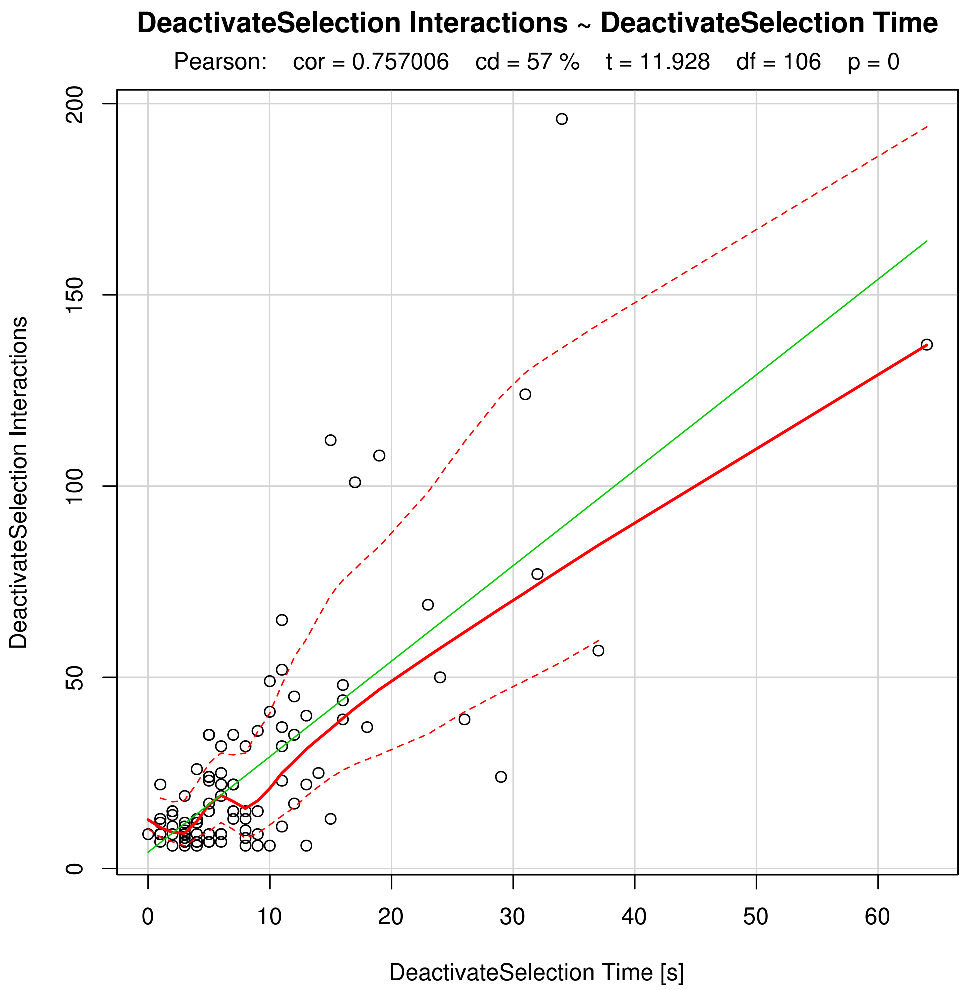


Figure 6.82: Insert Text Interactions in correlation to Time

This correlation shows that the more time was taken, the more interactions had been performed. These results confirm Hypothesis 6, which proposes that the more effective the user interaction design concept for inserting a new text is, the less time will be taken and the less interactions will be performed.

6.6 Conclusions

An investigation into the effectiveness of the proposed DOKY user interface has been carried out to see whether the proposed user interface design concepts and user interaction design concepts are effective means for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices, by automated structured observations of 876 blind and visually impaired research subjects performing 19 exercises among a highly structured example document using the DOKY Structured Observation App on their own mobile or wearable device remotely over the Internet.

The first 6 exercises focussed on the user interface design concepts and their hypotheses about the behaviour of blind and visually impaired persons among structured documents in relation to the reactions for presentation provided by the different output modalities which are the Earcon Output Modality, the Tacton Output Modality and the Speech Output Modality using earcons, tactons and synthetic speech.

The proposed user interface design concepts for presenting the element type of a document structure element to the user by the timbre of the single-pitch inherited elementary earcon, the rhythm of the single-motive inherited elementary tacton and the voice of the inherited elementary speech utterance showed to be effective.

The effectiveness of the proposed user interface design concepts for presenting the level of a document structure element to the user by the pitch of the single-pitch inherited elementary earcon, the tempo of the single-motive inherited elementary tacton and the pitch of the inherited elementary speech utterance depends on the hardware device used.

The effectiveness of the proposed user interface design concepts for presenting the position of a document structure element to the user by the spatial location of the single-pitch inherited elementary earcon, the position on the vertical y-axis of the multi-touch screen or multi-touch pad of the mobile and wearable device from the top to the bottom, the position on the horizontal orientation axis of the device anticlockwise within an angle of 180 and the spatial location of the inherited elementary speech utterance is depends on the input modality used.

As with the position, the effectiveness of the proposed user interface design concepts for presenting the length of a document structure element to the user by the length on the vertical y-axis of the multi-touch screen or multi-touch pad of the mobile and wearable device from the top to the bottom and the length on the horizontal orientation axis of the device anticlockwise within an angle of 180 degrees depends on the input modality used.

The proposed user interface design concepts for presenting the relationship of all nested document structure element or text nodes at a position to the user by rendering their single-pitch inherited elementary earcons simultaneously as parallel compound earcons, rendering their single-motive inherited elementary tactons sequentially as serial compound tactons and by rendering their inherited elementary speech utterances sequentially as serial compound utterances is partially effective.

As with the position and length, the proposed user interface design concepts for presenting the text of a document structure element or text node to the user by the text of the inherited elementary speech utterance is dependent on the input modality used.

The next 7 exercises focussed on the user interface design concepts and their hypotheses about the behaviour of blind and visually impaired persons among structured documents in relation to the actions for navigation provided by the different input modalities which are the Multitouch Input Modality and the Motion Input Modality.

The proposed user interface design concepts for setting the structure cursor by the user by putting down one pointer at a specific position or by rotating the device horizontally anticlockwise 10 degrees showed to be effective.

The effectiveness of the proposed user interface design concepts for moving the structure cursor in order to obtain a fast overview over the structure of the entire document by moving the one pointer up and down at an arbitrary speed or by rotating the device horizontally clockwise or anticlockwise within an angle of 180 degrees depends on the input modality used.

The effectiveness of the proposed user interface design concepts for setting the text cursor by the user by moving the one pointer more than 50 pixels to the left or to the right or by rotating the device vertically anticlockwise 10 degrees is effective is dependent on the use of screen reader.

As with moving the structure cursor, the effectiveness of the proposed user interface design concepts for moving the text cursor forward and backward in order to obtain a fast overview over the entire text contained in a document structure element by moving the one pointer left or right at an arbitrary speed or by rotating the device vertically clockwise or anticlockwise within an angle of 90 depends on the input modality used.

The effectiveness of the proposed user interface design concepts for unsetting the text cursor by moving the one pointer more than 50 pixels up or down or by rotating the device vertically back to the start position is effective is dependent on the hardware device used.

As with moving the text cursor, the effectiveness of the proposed user interface design concepts for unsetting the structure cursor by lifting up the one pointer or by rotating the device horizontally back to the start position depends on the input modality used.

Finally, the effectiveness of the proposed user interface design concepts for setting the select modifier by putting down an additional second pointer at an arbitrary position or by flipping the device 90 degrees to the left is dependent on the input modality used too.

The last 6 exercises focussed on the user interaction design concepts and their hypotheses about the behaviour of blind and visually impaired persons among structured documents in relation to the interactions for manipulation provided by the user interaction design.

The effectiveness of the proposed user interaction design concept for activating a selection of the document for manipulation by setting the active modifier by using one of the different input modalities depends on the input modality used.

As with activating a selection, the effectiveness of the proposed user interaction design concept for deactivating a selection of the document for manipulation by unsetting the active modifier by using one of the different input modalities is dependent on the input modality used too.

The proposed user interaction design concept for moving a selection of the document to another position within the document by focusing the target position within the document to where the selection should be moved and unsetting the active modifier by using one of the different input modalities showed to be effective.

The effectiveness of the the proposed user interaction design concept for removing a selection of the document from the document by unsetting the focus from the structure element which currently has the focus and unsetting the active modifier by using one of the different output modalities depends on the input modality used.

As with removing a selection, the effectiveness of the proposed user interaction design concept for inserting a new structure element into the document by setting the insertElement cursor to a position on the insertElement axis by using one of the different input modalities, optionally moving the insertElement cursor on the insertElement axis forward or backward to choose another element type and setting the select modifier by using one of the different input modalities is dependent on the input modality used.

Finally, the effectiveness of the proposed user interaction design concept for inserting a new text into the document by setting the insertText modifier by using one of the different input modalities, speaking the text to insert and unsetting the insertText modifier by using one of the different input modalities depends on the input modality used too.

Table 6.1 summarises the 19 user interface design and user interaction design concepts and their hypotheses about the behaviour of blind and visually impaired persons among structured documents in relation to the actions for navigation, reactions for presentation and interactions for manipulation:

Concept	Hypothesis					
	1	2	3	4	5	6
Presentation						
Element Type	Confirm	Confirm	Confirm	Reject	Confirm	Confirm
Level	Confirm	Confirm	Confirm	Reject	Reject	Confirm
Position	Confirm	Confirm	Reject	Reject	Confirm	Confirm
Length	Confirm	Confirm	Reject	Reject	Confirm	Confirm
Relationship	Confirm	Reject	Confirm	Confirm	Confirm	Confirm
Text	Confirm	Confirm	Reject	Reject	Confirm	Confirm
Navigation						
Set Structure Cursor	Confirm	Confirm	Confirm	Reject	Confirm	Reject
Move Structure Cursor	Confirm	Confirm	Reject	Reject	Confirm	Confirm
Set Text Cursor	Confirm	Confirm	Confirm	Confirm	Confirm	Confirm
Move Text Cursor	Confirm	Confirm	Reject	Reject	Confirm	Confirm
Unset Text Cursor	Confirm	Confirm	Confirm	Reject	Reject	Confirm
Uset Structure Cursor	Confirm	Confirm	Reject	Reject	Confirm	Confirm
Set Select Modifier	Confirm	Confirm	Reject	Reject	Confirm	Confirm
Manipulation						
Activate Selection	Confirm	Confirm	Reject	Reject	Confirm	Confirm
Deactivate Selection	Confirm	Confirm	Reject	Reject	Confirm	Confirm
Move Selection	Confirm	Confirm	Confirm	Reject	Confirm	Confirm
Remove Selection	Confirm	Confirm	Reject	Reject	Confirm	Confirm
Insert Element	Confirm	Confirm	Reject	Reject	Confirm	Confirm
Insert Text	Confirm	Confirm	Reject	Reject	Confirm	Confirm

Table 6.1: User Interface Design and Interaction Concepts and Hypotheses

7 Conclusions

This chapter concludes the thesis by summarising the achievements of the research programme and discussing the limitations of the performed research. In addition, considerations about possible directions of further research based on the presented research results within the area of non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices are proposed.

7.1 Achievements of the Research

Literature within related fields of research about what structured documents are in general was reviewed and the essential components involved in the process of non-visual reading, creating and editing of structured documents had been identified. The different approaches available for each component were described in greater detail and their potential applications in a user interface for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices had been discussed. In addition, some examples of existing systems that use each non-visual approach in their human computer user interfaces were provided.

A research methodology for developing the user interface by employing a natural science epistemological model (positivism), an objective ontology (objectivism), a quantitative research strategy and an iterative research approach has been proposed. The user interface has been created using an inductive research approach where general research questions stand at the beginning and the theory, in this case the user interface design and the user interaction design, is the outcome of the research. For the evaluation of the user interface, a deductive research approach has been followed where the purpose of the theory is to generate the hypotheses that can be tested and that will thereby allow explanations of laws to be assessed. The last step involved a movement that is in the opposite direction from deduction to induction, as the implications of the findings from the theory that prompted the whole exercise had been inferred. These findings were fed back into the stock of theory and the research findings associated with the domain of non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices.

In order to develop a novel concept for a multi-modal user interface for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices, a survey has been conducted among 205 blind and visually impaired persons to find out how they handle structured documents at the moment and what is of importance as to the reading of and the navigation within structured documents from the blind and visually impaired person’s point of view. The results showed that there is serious need for action required by conducting further research on a novel concept for an assistive technology in support of blind and visually impaired persons for the reading of structured documents. Further research should be investigated in the digital physical medium. A novel concept should be format independent. More than one human sense should be served. Novel input modalities like multi-touch gestures, motion gestures or speech should be employed. The logical document structure plays a very important role as to the reading of and the navigation within structured documents. At the moment most structured documents available on the Internet are not or not fully accessible for blind and visually impaired persons and novel concepts should be developed for and implemented on the emerging category of mobile and wearable devices like smart phones, smart tablets or smart watches.

A multi-modal user interface for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices like smart phones, smart watches or smart tablets has been developed as a result of the inductive research among 205 blind and visually impaired participants in the previous chapters. It enables the user to get a fast overview over the document structure and to efficiently skim and scan over the document content by identifying the type, level, position, length, relationship and content text of each element as well as to focus, select, activate, move, remove and insert structure elements or text. These interactions are presented in a non-visual way using Earcons, Tactons and synthetic speech utterances, serving the auditory and tactile human sense. Navigation and manipulation is provided by using the multitouch, motion (linear acceleration and rotation) or speech recognition input modality. It is a complete solution for reading, creating and editing structured documents in a non-visual way. There is no special hardware required.

The Node Array Document Model has been developed, which is an alternative modality-neutral document model for presentation, navigation and manipulation of structured documents. In contrast to the tree-like document representation employed in the Document Object Model (DOM), the element and text objects of the document are organised in a two-dimensional array. The advantages of this representation is that it enables the user to get a fast overview over the document structure and to efficiently skim and scan over the document content by identifying the type, level, position, length, relationship and text of each element.

The Earcon Output Modality has been developed, which presents the interactions of the node array document model using earcons. These are non-verbal audio messages that are used in the computer user interface to provide information to the user about some computer object, operation or interaction. Each interaction is presented by a single-pitch inherited elementary earcon inheriting the timbre of the element type, pitch of the element level, spatial location of the element position, register of the operation and dynamics of the status. These single-pitch inherited elementary earcons are rendered simultaneously as parallel compound earcons for all interactions that are going on at the same time to reduce the length of time a compound audio message takes.

The Tacton Output Modality has been developed, which presents the interactions of the node array document model using tactons or tactile icons. These are structured, abstract vibrotactile messages that can be used for presenting multidimensional information non-visually. Each interaction is presented by a single-motive inherited elementary tacton inheriting the rhythm of the element type, the tempo of the element level and the intensity of the operation. These single-motive inherited elementary tactons are rendered sequentially as serial compound tactons for all interactions that are going on at the same time.

The Speech Output Modality has been developed, which presents the interactions of the node array document model using synthetic speech. Each interaction is presented by a inherited elementary utterance, inheriting the voice of the element type, pitch of the node level, spatial location of the node position, register of the operation and text of the content. These inherited elementary utterances are rendered sequentially as serial compound utterances for all interactions that are going on at the same time.

The Multitouch Input Modality has been developed, which allows the user to perform the actions provided by the node array document model by performing multi-touch gestures on the multi-touch screen or multi-touch pad of the mobile and wearable device. This input modality is mainly purposed for smart-phones, smart-pablets and smart-tablets because these category of devices typically have a large multi-touch screen embedded and may be too large and heavy to be moved around themselves for using motion.

The Orientation Motion Input Modality has been developed, which allows the user to perform the actions provided by the node array document model by performing motion gestures by moving the mobile or wearable device itself. This input modality is mainly purposed for wearable devices like smart-watches because these category of devices typically are very small and lightweight and can therefore be moved around easily. In addition there may be only a very small multi-touch screen or multi-touch pad or no multi-touch screen or multi-touch pad at all embedded in these devices making the use of multi-touch input difficult.

A flexible platform-independent and event-driven software architecture implementing the DOKY user interface as well as the automated structured observation research method employed for the evaluation of the effectiveness of the proposed user interface has been presented. Because it is platform- and language-neutral, it can be used in a wide variety of platforms, environments and applications for mobile and wearable devices. Each component is defined by interfaces and abstract classes only, so that it can be easily changed or extended, and grouped in a semantically self-containing package.

An investigation into the effectiveness of the proposed DOKY user interface has been carried out to see whether the proposed user interface design concepts and user interaction design concepts are effective means for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices, by automated structured observations of 876 blind and visually impaired research subjects performing 19 exercises among a highly structured example document using the DOKY Structured Observation App on their own mobile or wearable device remotely over the Internet. The results showed that the proposed user interface design concepts for presentation and navigation and the user interaction design concepts for manipulation are effective and that their effectiveness depends on the input modality and hardware device employed as well as on the use of screen readers.

The research has met all of the aims and objectives originally outlined in Chapter 1.1 and has resulted in the design and development of a multi-modal user interface for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices. In addition, a number of papers relating to the research programme have been published in internationally recognised journals and presented at refereed conferences. A list of the publications can be found in the Appendix D. As such, it is believed that the research has made valid and useful contributions to the field of non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices.

7.2 Limitations of the Research

Although the objectives of the research project have been met, a number of decisions had to be made which imposed limitations upon the work. These decisions were typically either practically based or due to financial restrictions. The key limitations of the research are summarised below.

The participating countries and organisations had been selected randomly using simple random sampling as probability sampling method. This probability sampling method is likely to generate a representative sample which reflects the population accurately so that it is a microcosm of the population from which it has been taken and sampling error is kept to a minimum. Therefore external validity is high and findings can be generalised to the population because each unit in the population has the same chance of being selected. However, It is extremely unlikely that one will end up with a truly representative sample, even when probability sampling is employed because of different sources of sampling error and non-sampling error.

Sampling error is the error in the findings deriving from research due to the difference between a sample and the population from which it is selected. This may occur even though probability sampling has been employed. In this research, some sampling error derive from the fact that there is a large dark number of blind and visually impaired persons who are not member of any organisation and therefore they are unreachable for this research.

Non-sampling error is the error in the findings deriving from research due to the difference between the population and the sample that arise from deficiencies in the sampling approach. A source of non-sampling error in this research is non-response. This occurred when some members of the sample refused to cooperate. For example if an organisation did not forward the invitation letter to its members or a member did not respond to the invitation letter, could not be contacted or could not supply the required data for example, because of mental incapacity. Invitations sent out over the Internet are prone to poor response rates.

The limitation of the non-probability sampling methods employed for sampling the blind and visually impaired research subjects of all organisations in each selected country is that the resulting sample may not be a representative sample of the population and hence it may not be possible to generalize the findings to the population although the results are statistically significant because the research subjects had not been selected randomly as with a probability sampling method. This implies that some units in the population are more likely to be selected than others. For example in case of the convenience sample they all are members of the same organisations or in case of the snowball sample they all know each other. Therefore external validity is typically low.

However these non-probability sampling methods were the only way possible to sample blind and visually impaired subjects for this research because there is no directory of all blind and visually impaired persons in the world and the organisations are not allowed to give out the records of their members for data protection reasons. The organisations can only send out an invitation letter to all of their members and if a member is interested in participating in this research he or she can get in contact by responding to this invitation letter. Therefore it was not possible to actively select the research subjects employing probability sampling methods like simple random sampling as used for the selection of the participating countries. Instead the sample was passively formed of all respondents who replied to the invitation letter.

Wearable devices were not tested and motion interaction was only tested to a limited degree since nobody of the persons concerned used a smart watch because this category of hardware device is very new and therefore not already employed by the respondents.

Despite these limitations, the research project has made valid contributions to knowledge on non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices and provided sufficient proof of concept for the user interface design concepts and user interaction design concepts proposed.

7.3 Directions of Further Research

This research project has advanced the field of non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices. However, a number of areas of scope for future work can be identified, which are based on the results of this research project. Some of these ideas have already been mentioned in previous chapters. These suggestions for future work are detailed below:

Further resources could be spent into optimising the Orientation Motion Input Modality because this input modality has shown to be less effective than the Multitouch Input Modality in many of the proposed user interface design concepts for presentation and navigation and user interaction design concepts for manipulation since more time had to be taken and more interactions had to be performed by the research subjects to perform a navigation action, recognise an element attribute or text in question or to perform a manipulation interaction. In addition, document structure elements containing only a small portion of text may appear too small on the clock to be focused and selected reliable because there is no minimal unit size with the Orientation Motion Input Modality. To solve this problem, a minimal unit size which can be reliable focused should be introduced. Further research in the domain of motion gestures and psychophysics is required to find out how big this minimal unit size should be.

Additional effort could be put into the proposed user interface design concepts for presenting the relationship of all nested document structure element or text nodes at a position to the user, which had shown to be only partially effective. The relationship is presented to the user by rendering their single-pitch inherited elementary earcons simultaneously as parallel compound earcons to reduce the length of time this compound audio message takes, rendering their single-motive inherited elementary tactons sequentially as serial compound tactons and by rendering their inherited elementary speech utterances sequentially as serial compound utterances.

These results fit in well with the empirically derived guidelines for the presentation of concurrent earcons of McGookin & Brewster (2004) where the identification of earcons fell from 70 per cent, consistent with similar work by Brewster (1994) on single earcon identification, when only one earcon was presented, to 30 per cent when four earcons were concurrently presented (McGookin & Brewster, 2003). Identification of both timbre and rhythm encoded data attributes is much higher, dropping from 95 per cent to 65 per cent as the number of concurrently presented earcons is increased. Identification of the register encoded earcon attribute, whilst being much lower than timbre and rhythm, does fall at a much shallower gradient, due in part to the incorporation of inharmonic musical intervals between the registers used (McGookin & Brewster, 2004).

In addition, some research subjects complained that they feel annoyed by the continuous sounds. In this cases, the continuous sounds failed to fade into the background of consciousness after a short period of time as proposed by auditory habituation (Buxton, Gaver & Bly, 1991), which is one important ability of the human auditory system, and come to the foreground of attention if the sound was to change or stop because of the sensitivity of the auditory system to changes in stimulus (Buxton, 1989).

The serial compound tactons might not be able to keep up with the pace of all interactions that are going on in the user interface at the same time because they can take a long time to play since each motive lasts a particular length of time depending on its notes and the tempo and these are then combined to produce longer compound tactons. They could be played more rapidly (at a faster tempo) to overcome this problem but then errors in recognition may occur.

As an alternative approach to complex musical instrument timbres, spearcons or speech based earcons (Walker, Nance & Lindsay, 2006) may be employed for presenting the element type to the user. These are brief audio cues that can play the same roles as earcons and auditory icons, but in a more effective manner. Spearcons had been discussed in detail in Chapter 2.2.3. The advantages of spearcons are that they can be created automatically by converting the text of the name of an element type, for example “Heading”, to speech via text-to-speech (TTS), and then speeding up the resulting audio clip (without changing pitch) to the point that it is no longer comprehensible as speech. Spearcons are unique to the specific element type. These unique sounds are analogous to fingerprints because of the acoustic relation between the spearcons and the original speech phrases. At the same time, the similarities in element types cause the spearcons to form families of sounds. For example, the spearcons for “Table”, “Table Row” and “Table Cell” are all unique, including being of different lengths. However, they are acoustically similar at the beginning of the sounds, which allows them to be grouped together even though they are not comprehensible as any particular words.

Since the mapping between spearcons and their element types is non-arbitrary, there is less learning required than would be the case for a purely arbitrary mapping as with the complex musical instrument timbres. Spearcons provide more direct mappings between sound and element type than musical instrument timbres, and cover more content domains, more flexibly. Spearcons can be created algorithmically, so they can be created dynamically, and can represent any possible element type. Also, spearcons are easy to learn whether they are comprehensible as a particular word or not, because they derive from the original speech (Palladino & Walker, 2007). This has not been tested in this research. A further test would be required to investigate if spearcons are an effective method for presenting the element type to the user.

In addition, the element type of the element node associated with an interaction and on which the operation of the interaction is performed could be presented by the voice of the inherited elementary speech utterance. Sorin, Lemarié, Aussenac-Gilles, Mojahid & Oriola (2014) used specialized audio and two voices to demarcate headings in their work on communicating text structure to blind persons with text-to-speech. Their research showed that text comprehension was slightly improved but failed to show statistically significant evidences that either the dual-voices method or the spatialised audio method has better performance than current text-to-speech text description saying “Heading level N” before the heading oralisation. However, their document structures were very simple, only consisting of two different element types, headings and non-heading document contents. If more complex document structures were used then a performance increase may have been found. This has not been tested in this research. A further test would be required to investigate if different voices are an effective method for presenting the element type to the user.

In addition to the highly structured example text-document employed in the investigation into the effectiveness of the proposed DOKY user interface as presented in Chapter 6, further automated structured observations of research subjects performing exercises among other content types of digital structured documents available on the Internet, for example text-documents, spreadsheets, presentations, drawings, mathematical formulae or equations and web-sites, using the DOKY Structured Observation App on their own mobile or wearable devices remotely over the Internet, can be conducted.

In addition to the DOKY Structured Observation App, which is an application of all the different components for conducting automated structured observations of research subjects performing exercises among a highly structured example document on their own mobile or wearable devices remotely over the Internet, other applications of the DOKY user interface in the form of user agents or authoring tools can be developed. For example, a non-visual Web-Browser App or a non-visual Word-Processor App can provide a complete solution for reading, creating, editing, loading and saving real structured documents in a non-visual way.

Future work might also have the potential for a longitudinal study, involving participants adopting the day-to-day use of DOKY, and providing both quantitative and qualitative feedback.

8 References

8.1 Literature

1. Adelstein, B.D., Begault, D.R., Anderson, M.R., and Wenzel, E.M. Sensitivity to Haptic-Audio Asynchrony. Proc. 5th International Conference on Multimodal Interfaces, ACM Press (2003), 73-76.
2. Aldrich, F.K. & Parkin, A.J. (1989). Listening at speed. British journal of visual impairment and blindness, 7(1), pp. 16-18.
3. Allan, J., Lowney, G., Patch, K., Spellman, J., 2015. W3C WAI User Agent Accessibility Guidelines (UAAG) 2.0. http://www.w3.org/TR/UAAG20/, visited November 2018
4. Android Developers, 2010. Android Access Framework Reference http://developer.android.com/reference/android/view/accessibility/package-summary.html, visited March 2011
5. Apple, 2009. UIAccessibility Protocol Reference. http://developer.apple.com/library/ios/#documentation/uikit/reference/UIAccessibility_Protocol, visited June 2010
6. Asakawa, C., Takagi, H.: Maximum Listening Speed For The Blind. In: Proceedings of the 2003 International Conference on Auditory Display, pp. 276-279 (2003).
7. Badeley, A. (1990). Human Memory: Theory and Practice. London: Lawrence Erlbaum Associates.
8. Ballas, J.A. & Howard, J.H. (1987). Interpreting the language of environmental sounds. Envirionment and Behaviour, 19(1), pp. 91-114.
9. Barker, P.G. & Manji, K.A. (1989). Pictorial dialogue methods. International Journal of Man-Machine Studies, 31, pp. 323-347.
10. Bianchetti, R., Erle, M., Hofer, S., 2012. Mainstreaming the Creation of Accessible PDF Documents by a Rule-Based Transformation from Word to PDF. In Proceedings of the 12th International Conference on Computers Helping People with Special Needs (Linz, Austria, July 11 – 13, 2012). ICCHP 2012.
11. Björk, E.A. (1985). The perceived quality of natural sounds. Acustica, 57(3), pp. 185-188.
12. Blattner, M., Sumikawa, D. & Greenberg, R. (1989). Earcons and icons: Their structure and common design principles. Human Computer Interaction, 4(1), pp. 11-44.
13. Blattner, M., Papp, A. & Glinert, E. (1992). Sonic enhancements of two-dimensional graphic displays. In G. Kramer (Ed.), Auditory Display, sonification, audification and auditory interfaces. The Proceedings of the First International Conference on Auditory Display, Santa Fé Institute, Santa Fé: Addison-Wesley, pp. 447-470.
14. Brewster, S.A. (1992). Providing a model for the use of sound in user interfaces. University of York Technical Report YCS 169, York, UK.
15. Brewster, S.A., Wright, P.C. & Edwards, A.D.N. (1992). A detailed investigation into the effectiveness of earcons. In G. Kramer (Ed.), Auditory display, sonification,

audification and auditory interfaces. The Proceedings of the First International Conference on Auditory Display, Santa Fé Institute, Santa Fé: Addison-Wesley, pp. 471-498.

16. Brewster, S.A., Wright, P.C. & Edwards, A.D.N. (1993). An evaluation of earcons for use in auditory human-computer interfaces. In S. Ashlund, K. Mullet, A. Henderson, E. Hollnagel, & T. White (Ed.), INTERCHI'93, Amsterdam: ACM Press, Addison-Wesley, pp. 222-227.
17. Brewster, S.A., Wright, P.C. & Edwards, A.D.N. (1993). Parallel earcons: Reducing the length of audio messages. Submitted to the International Journal of Man-Machine Studies.
18. Brewster, S.A., Providing a Structured Method for Integrating Non-Speech Audio into Human-Computer Interfaces, PhD Thesis, Department of Computer Science, University of York, 1994.
19. Brewster, S.A., Wright, P.C. & Edwards, A.D.N. Experimentally Derived Guidelines for the Creation of earcons. In Proc. HCI '95, Springer (1995), 155 - 159.
20. Brewster, S.A. (1998): Using Non-Speech Sounds to Provide Navigation Cues. ACM Transactions on Computer-Human Interaction, 5, 224-259.
21. Brewster, S.A. and Brown, L.M. (2004) Tactons: structured tactile messages for non-visual information display. In, Australasian User Interface Conference 2004, ACS Conferences in Research and Practice in Information Technology Vol 28, pages pp. 15-23, Dunedin, New Zealand.
22. Brewster, S.A., Brown, L.M., 2004. Tactons: structured tactile messages for non-visual information display. In Proceedings of the fifth conference on Australasian user interface - Volume 28, Pages 15-23 (Dunedin, New Zealand, January 18 – 22, 2004). AUIC 2004.
23. Brown, M.L., Newsome, S.L. & Glinert, E.P. (1989). An experiment into the use of auditory cues to reduce visual workload. In Proceedings of CHI'89, Austin, Texas: ACM Press, Addison-Wesley, pp. 339-346.
24. Brown, L. M., Brewster, S. A., and Purchase, H. C. A First Investigation into the Effectiveness of tactons. In Proc. WorldHaptics 2005, IEEE (2005), 167 – 176.
25. Brown, L. M. and Brewster, S. A. Multidimensional tactons for Non-Visual Information Display in Mobile Devices. In Proc. MobileHCI '06, ACM Press (2006), 231 – 238.
26. Brown, L. M., Brewster, S. A., and Purchase, H. C. Tactile crescendos and sforzandos: applying musical techniques to tactile icon design. In Proc. Vol II CHI 2006, ACM Press (2006), 610 - 615.
27. Brown, L.M., "Tactons: Structured Vibrotactile Messages for Non-Visual Information Display", PhD Thesis, Department of Computing Science, University of Glasgow, 2007.
28. Bruce, I. A. and McKennell, A. Blind and Partially Sighted Adults in Britain: The RNIB Survey. HMSO, London, 1991.
29. Brugger, R.: Eine statistische Methode zur Erkennung von Dokumentstrukturen. PhD thesis no. 1251, University of Fribourg, 1998
30. Bryman, A., & Cramer, D. (2011). Quantitative data analysis with IBM SPSS 17, 18 and 19. Routledge.

31. Bryman, A. (2016). Social research methods. Oxford university press.
32. Burgess, D. (1992). Techniques for low cost spatial audio (Technical Report No. GIT-GVU-92-09). Graphics, Visualization & Usability Centre, Georgia Institute of Technology.
33. Buxton, W. (1989). Introduction to this special issue on nonspeech audio. Human Computer Interaction, 4(1), pp. 1-9.
34. Buxton, W., Gaver, W. & Bly, S. (1991). Tutorial number 8: The use of non-speech audio at the interface. In Proceedings of CHI'91, New Orleans: ACM Press: Addison-Wesley.
35. Chan, A., "Designing Haptic Icons to Support an Urgency-Based Turn-Taking Protocol", M.Sc. Thesis, Computing Science, University of British Columbia, 2004.
36. Chan, A., MacLean, K.E., and McGrenere, J., "Learning and Identifying Haptic Icons under Workload", in Proceedings of World Haptics 2005, IEEE Press (2005), pp 432-439.
37. Chowning, J. (1975). Synthesis of complex audio spectra by means of frequency modulation. Journal of the Audio Engineering Society, 21(7), pp. 526-534.
38. Cohen, J. (1992). Monitoring background activities. In G. Kramer (Ed.), Auditory Display, sonification, audification and auditory interfaces. The Proceedings of the First International Conference on Auditory Display., Santa Fé Institute, Santa Fé: Addison-Wesley, pp. 499-532.
39. Cohen, J. (1993). "Kirk Here": Using genre sounds to monitor background activity. In S. Ashlund, K. Mullet, A. Henderson, E. Hollnagel, & T. White (Eds.), INTERCHI'93, Adjunct Proceedings Amsterdam: ACM Press, Addison Wesley, pp. 63-64.
40. Colquhoun, W.P. (1975). Evaluation of auditory, visual and dual-mode displays for prolonged sonar monitoring tasks. Human Factors, 17, pp. 425-437.
41. Cooper, M., Kirkpatrick, A., O'Connor, J., 2016. Techniques and Failures for Web Content Accessibility Guidelines 2.0. http://www.w3.org/TR/WCAG20-TECHS, visited November 2018
42. Craig, J. C. and Sherrick, C. E. (1982): Dynamic Tactile Displays. In Tactual Perception: A Sourcebook (Ed, Foulke, E.) Cambridge University Press, pp. 209-233.
43. Davishy, A., Hutter, H.P., Horvath, A., Dorigo, M.: A Flexible Software Architecture Concept for the Creation of Accessible PDF Documents. In proceedings of the 12th International Conference on Computers Helping People with Special Needs (Wien, Austria, July 14 – 16, 2010), LNCS 6179, Springer Verlag Berlin. ICCHP 2010.
44. Darvishy, A., Hutter, H.P., Mannhart, O., 2011. Web Application for Analysis, Manipulation and Generation of Accessible PDF Documents. In Proceedings of the 14th International Conference on Human-Computer Interaction (Orland, Florida, USA, July 09 – 14, 2011). HCII 2011.
45. Darvishy, A., 2018. PDF Accessibility: Tools and Challenges. In International Conference on Computers Helping People with Special Needs (pp. 113-116). Springer, Cham.

46. Darvishy, A., Hutter, H. P., 2013. Comparison of the effectiveness of different accessibility plugins based on important accessibility criteria. In International Conference on Universal Access in Human-Computer Interaction (pp. 305-310). Springer, Berlin, Heidelberg.

47. Darvishy, A., Nevill, M., Hutter, H. P., 2016. Automatic Paragraph Detection for Accessible PDF Documents. In International Conference on Computers Helping People with Special Needs (pp. 367-372). Springer, Cham.

48. Darvishy, A., Leemann, T., Hutter, H.P., 2012. Two Software Plugins for the Creation of Fully Accessible PDF Documents Based on a Flexible Software Architecture. In Proceedings of the 12th International Conference on Computers Helping People with Special Needs (Linz, Austria, July 11 – 13, 2012). ICCHP 2012.

49. Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. Perception and Psychophysics, 28(5), pp. 381-389.

50. Deutsch, D. (1986). Auditory pattern recognition. In K. R. Boff, L. Kaufman, & P. Thomas (Eds.), Handbook of perception and human performance, pp. 32.1 - 32.49. New York: Wiley.

51. Dewar, K.M., Cuddy, L.L. & Mewhort, D.J. (1977). Recognition of single tones with and without context. Journal of Experimental Psychology: Human Learning and Memory, 3(1), pp. 60-67.

52. Diggs, J., McCarron, S., Cooper, M., Schwerdtfeger, R., Craig, J., 2017. Accessible Rich Internet Applications (WAI-ARIA) 1.1. http://www.w3.org/TR/wai-aria-1.1/, visited November 2018

53. Dillman, D. A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J., Messer, B. L. (2009). Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. Social science research, 38(1), 1-18.

54. Donker, H., Klante, P., & Gorny, P., “The Design of Auditory User Interfaces for Blind Users,” presented at NordiCHI, October, 2002, available at http://www-cg-hci.informatik.uni-oldenburg.de/resources/zib_nordichi.pdf.

55. Dov Dori, David Doermann, Christian Shin, Robert Haralick, Ishin Phillips, Mitchell Buchmann, and David Ross. The representation of document structure: A generic object-process analysis. In H. Bunke and P. S. P. Wang, editors, Handbook of Optical Character Recognition and Document Analysis. World Scientific, 1997.

56. ECMA-376 Standard. (2006). Office Open XML File Formats. http://www.ecma-international.org/publications/standards/Ecma-376.htm, visited June 2012

57. ECMA-388 Standard. (2009). Open XML Paper Specification. http://www.ecma-international.org/publications/standards/Ecma-388.htm, visited June 2012

58. Enriquez, M. J. and MacLean, K. (2003): The Hapticon Editor: A tool in support of haptic communication research. Haptics Symposium 2003, Los Angeles, CA, 356-362, IEEE Press

59. Enriquez, M., MacLean, K., and Chita, C. Haptic Phenomes: Basic Building Blocks of Haptic Communication. In Proc. 8th International Conference on Multimodal Interfaces (ICMI'06), IEEE (2006), 302 - 309.

60. Faulkner, S., Eicholz, A., Leithead, T., Danilo, A., Moon, S. (2017). HTML 5.2 Specification. http://www.w3.org/TR/html52/, visited November 2018

61. Foulke, E., "Reading Braille", in Tactual Perception: A Sourcebook, W. Schiff and E. Foulke (ed). Cambridge University Press, 1982, pp 168-208.

62. Frysinger, S.P. (1990). Applied research in auditory data representation. In D. Farrell (Ed.), Extracting meaning from complex data: processing, display, interaction. Proceedings of the SPIE/SPSE symposium on electronic imaging. 1259 Springfield, VA.: SPIE, pp. 130-139.

63. Gemperle, F., Ota, N., and Siewiorek, D., "Design of a Wearable Tactile Display", in Proceedings of the Fifth International Symposium on Wearable Computers, (2001), pp 5-12.

64. Gaver, W. (1986). Auditory Icons: Using sound in computer interfaces. Human Computer Interaction, 2(2), pp. 167-177.

65. Gaver, W. (1989). The SonicFinder: An interface that uses auditory icons. Human Computer Interaction, 4(1), pp. 67-94.

66. Gaver, W. (1992). Using and creating auditory icons. In G. Kramer (Ed.), Auditory Display, sonification, audification and auditory interfaces. The Proceedings of the First International Conference on Auditory Display, Santa Fé Institute, Santa Fé: Addison-Wesley, pp. 417-446.

67. Gaver, W. (1993). Synthesizing auditory icons. In S. Ashlund, K. Mullet, A. Henderson, E. Hollnagel, & T. White (Ed.), INTERCHI'93, Amsterdam: ACM Press, Addison-Wesley, pp. 228-235.

68. Gaver, W. & Smith, R. (1990). Auditory icons in large-scale collaborative environments. In D. Diaper, D. Gilmore, G. Cockton, & B. Shackel (Eds.), Human Computer Interaction: Interact'90, Cambridge, UK: Elsevier Science Publishers B.V. (North Holland), pp. 735-740.

69. Gaver, W., Smith, R. & O'Shea, T. (1991). Effective sounds in complex systems: The ARKola simulation. In S. Robertson, G. Olson, & J. Olson (Eds.), Proceedings of CHI'91, New Orleans: ACM Press, Addison-Wesley, pp. 85-90.

70. Gaver, W., Moran, T., MacLean, A., Lövstrand, L., Dourish, P., Carter, K. & Buxton, W. (1992). Realizing a video environment: EuroParc's RAVE system. In P. Bauersfeld, J. Bennett, & G. Lynch (Eds.), Proceedings of CHI'92, Monterey, California: ACM Press, Addison-Wesley, pp. 27-35.

71. Gerhing, B. & Morgan, D. (1990). Applications of Binaural Sound in the Cockpit. Speech Technology, 5(2), pp. 46-50.

72. Geldard, F.A., "Adventures in tactile literacy." The American Psychologist 12, 1957, pp 115–124.

73. Gill, J. (2003), Vol. 2003 Royal National Institute of the Blind, United Kingdom.

74. Gibson, J. J. The Senses Considered as Perceptual Systems. Houghton Mifflin, Boston, 1966.

75. Gnome Accessibility Project, 2002. Assistive Technology Service Provider Interface AT-SPI. http://projects.gnome.org/accessibility, visited June 2010

76. Gunther, E. (2001): Skinscape: A Tool for Composition in the Tactile Modality. Massachusetts Institute of Technology. Masters of Engineering.

77. Gunther, E., Davenport, G. and O'Modhrain, S. (2002): Cutaneous Grooves: Composing for the Sense of Touch. Proceedings of Conference on New Instruments for Musical Expression, Dublin, IR, 1-6.

78. Hapeshi, K. & Jones, D. (1992). Interactive multimedia for instruction: A cognitive analysis of the role of audition and vision. International Journal of Human-Computer Interaction, 4(1), pp. 79-99.

79. Henry, S.L., Duffy, M., 2005. W3C WAI Essential Components of Web Accessibility http://www.w3.org/WAI/intro/components.php, visited February 2011

80. Ho, C., Spence, C., and Tan, H.Z., "Warning Signals Go Multisensory", in Proceedings of HCI International 2005, Lawrence Erlbaum Associates (2005), pp 1-10.

81. Ho, C., Tan, H.Z., and Spence, C., "Using spatial vibrotactile cues to direct visual attention in driving scenes", Transportation Research Part F: Traffic Psychology and Behaviour 8, 2005 pp 397-412.

82. Hoggan, E. and Brewster, S. A., Crossmodal Icons for Information Display, in Proc ACM CHI '06 Extended Abstracts, ACM Press (2006), 857 – 862.

83. Hoggan, E. and Brewster, S. A. Crossmodal Spatial Location: Initial Experiments. In Proc. NordiCHI '06, ACM Press (2006), 469 - 472.

84. Hoggan, E. and Brewster, S. A. Designing audio and tactile crossmodal icons for mobile devices. In Proc. Proceedings of the 9th international conference on Multimodal interfaces, ACM (2007), 162 – 169.

85. Hoggan, E. and Brewster, S. A. New Parameters for tacton Design. In Proc. Vol II CHI 2007, ACM Press (2007), 2417 - 2422.

86. Hoggan, E. and Brewster, S. A. CrossTrainer: Testing The Use of Multimodal Interfaces in Situ. In Proc. CHI'10, ACM Press (2010).

87. ISO/IEC 26300. (2006). Information technology - Open Document Format for Office Applications (OpenDocument) v1.0. http://www.iso.org/standard/43485.html, visited June 2012

88. ISO/IEC DTR 13066. (2007). Information technology - Interoperability with assistive technology (AT) - IAccessible2 accessibility application programming interface (API) http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=53998, visited March 2010

89. ISO/IEC 29500. (2008). Information technology - Document description and processing languages - Office Open XML File Formats. http://www.iso.org/standard/51463.html, visited June 2012

90. ISO 32000. (2008). Document management - Portable document format - PDF 1.7 http://www.iso.org/iso/catalogue_detail?csnumber=51502, visited March 2010

91. ISO 14289. (2012). Document management applications - Electronic document file format enhancement for accessibility - Use of ISO 32000 (PDF/UA) http://www.iso.org/iso/catalogue_detail.htm?csnumber=54564, visited June 2012

92. James, F., 1996. Presenting HTML Structure in Audio: User Satisfaction with Audio Hypertext. In Proceedings of the International Conference on Auditory Display (Palo Alto, CA, USA, November 4 - 6, 1996). ICAD '96

93. Jay, C., Stevens, R., Hubbold, R., Glencross, R., 2008. Using haptic cues to aid nonvisual structure recognition. In Journal ACM Transactions on Applied Perception (TAP) Volume 5, Issue 2, Article No. 8, May 2008, ACM New York, New York, USA.

94. Kildal, J. and Brewster, S.A., Explore the Matrix: Browsing Numerical Data Tables Using Sound. In Proc. ICAD2005, (2005).

95. Kildal, J., Brewster, S.A., 2006. Non-visual overviews of complex data sets. In Proceedings of the ACM SIGCHI conference on human factors in computing systems, Pages 947-952, (Monréal, Québec, Canada, April 22 – 27, 2006). CHI 2006.

96. Kildal, J., Brewster, S.A., 2006. Providing a size-independent overview of non-visual tables. In Proceedings of the 12th International Conference on Auditory Display, London, United Kingdom, June 20 - 23, 2006

97. Kirkpatrick, A., O'Connor, J., Campbell, A., Cooper, M., 2015. W3C WAI Web Content Accessibility Guidelines (WCAG) 2.1. http://www.w3.org/TR/WCAG21/, visited November 2018

98. Kramer, G. (1992). Some organizing principles for representing data with sound. In G. Kramer (Ed.), Auditory Display, sonification, audification and auditory interfaces. The Proceedings of the First International Conference on Auditory Display, Santa Fé Institute, Santa Fé: Addison-Wesley, pp. 185-222.

99. Lazar, J., Allen, A., Kleinman, A., Malarkey, C., 2007. What Frustrates Screen Reader Users on the Web: A Study of 100 Blind Users. In International Journal of Human-Computer Interaction Volume 22, Issue 3, Pages 247 – 269, 2007.

100. Levitt, H. & Voroba, B. (1974). Localization. In S. E. Gerber (Ed.), Introductory Hearing Science: physical and physiological concepts, pp. 188-196. Philadelphia: W.B. Saunders Company.

101. Lewkowicz, D. J., The Development of Intersensory Temporal Perception: An Epigenetic Systems/Limitations

102. MacLean, K. and Enriquez, M. Perceptual Design of Haptic Icons. In Proc. Eurohaptics, (2003), 351 - 363.

103. Marks, L. E. The Unity of Senses. Interrelations Among The Modalities. Academic Press, New York, 1978.

104. McGookin, D.K., "The Presentation of Multiple Earcons in a Spatialised Audio Space", in Volume II Proceedings of British HCI 2002, BCS (2002), pp 228-229.

105. McGookin, D.K. and Brewster, S.A. An investigation into the identification of concurrently presented earcons. In Proceedings of ICAD 2003, pages 42–46, Boston, Massachusetts, 2003. ICAD.

106. McGookin, D.K. and Brewster, S.A. Empirically derived guidelines for the presentation of concurrent earcons. In Proceedings of BCS-HCI 2004, volume 2, pages 65–68, Leeds, UK, 2004. BCS.

107. McGookin, D.K. and Brewster, S.A., "Space, The Final Frontearcon: The Identification Of Concurrently Presented Earcons In A Synthetic Spatialised Auditory Environment", in Proceedings of ICAD 2004, ICAD (2004), pp 1-8.

108. McGookin, D.K. and Brewster, S.A. Understanding Concurrent earcons: Applying Auditory Scene Analysis Principles to Concurrent earcon Recognition. ACM Transactions on Applied Perception 1 (2004), 130 - 155.

109. McGookin, D.K., "Understanding and Improving the Identification of Concurrently Presented Earcons", PhD Thesis, Department of Computing Science, University of Glasgow, 2004.

110. Microsoft, 1997. Microsoft Active Accessibility API MSAA. http://msdn.microsoft.com/en-us/library/dd373592%28VS.85%29.aspx, visited May 2011

111. Microsoft, 2005. Microsoft User Interface Automation API UIA. http://msdn.microsoft.com/en-us/library/ms747327.aspx, visited June 2010

112. Minatani, K., 2012. Development of a DAISY Player that Utilizes a Braille Display for Document Structure Presentation and Navigation. In Proceeding of the 12th International Conference on Computers Helping People with Special Needs (Linz, Austria, July 11 – 13, 2012). ICCHP 2012.

113. Monk, A. (1986). Mode Errors: A user-centered analysis and some preventative measures using keying-contingent sound. International Journal of Man-Machine Studies, 24, pp. 313-327.

114. Moore, B.C. (1989). An Introduction to the Psychology of Hearing (2nd ed.). London: Academic Press.

115. Mountford, S.J. & Gaver, W. (1990). Talking and listening to computers. In B. Laurel (Ed.), The art of human-computer interface design, pp. 319-334. Reading, Massachusetts: Addison-Wesley.

116. NISO Z39.86-2005 (R2012) Revision of ANSI/NISO Z39.86-2002, 2002, Specifications for the Digital Talking Book DAISY http://www.daisy.org/z3986/2005/Z3986-2005.html, visited March 2010

117. Palladino, D. K., & Walker, B. N. (2007). Learning Rates for Auditory Menus Enhanced with Spearcons versus Earcons. Proceedings of the 13th International Conference on Auditory Display, Montreal, pp. 274-279.

118. Palladino, D. K., & Walker, B. N. (2008). Efficiency of spearcon-enhanced navigation of one dimensional electronic menus. Proceedings of the 14th International Conference on Auditory Display, Paris, France.

119. Palladino, D. K., & Walker, B.N. (2008). Navigation efficiency of two dimensional auditory menus using spearcon enhancements, Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society (HFES 2008), San Antonio, Texas.

120. Patterson, R.D. (1982). Guidelines for auditory warning systems on civil aircraft (CAA Paper No. 82017). Civil Aviation Authority, London.

121. Perrott, D., Sadralobadi, T., Saberi, K. & Strybel, T. (1991). Aurally aided visual search in the central visual field: Effects of visual load and visual enhancement of the target. Human Factors, 33(4), pp. 389-400.

122. Petit, G., Dufresne, A., Levesque, V., Hayward, V., Trudeau, N.: Refreshable tactile graphics applied to schoolbook illustrations for students with visual impairment. Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility, pp. 89-96. ACM, Halifax, NS, Canada (2008)

123. Petit, G., Dufresne, A., Robert, J.M., 2011. Introducing TactoWeb: A Tool to Spatially Explore Web Pages for Users with Visual Impairment. In Proceedings of the 14th International Conference on Human-Computer Interaction (Orland, Florida, USA, July 09 – 14, 2011). HCII 2011.

124. Petrucci, L.S., Harth, E., Roth, P., Assimacopoulos, A., Pun, T. “WebSound: a generic Web sonification tool, and its application to an auditory Web browser for blind and visually impaired users”. In: Proceedings of the Sixth International Conference on Auditory Display (ICAD 2000). 2000.

125. Pierce, J., "Hearing in Time and Space", in Music, Cognition and Computerized Sound, P. R. Cook, Ed.: MIT Press, 2001, pp. 89-103.

126. Pitt, I. & Edwards, A. (1991). Navigating the interface by sound for blind users. In D. Diaper & N. Hammond (Eds.), Proceedings of HCI’91, VI Edinburgh: Cambridge University Press, pp. 373-383.

127. Pontelli, E., Gillan, D., Gupta, G., Karshmer, A., Saad, E., Xiong, W, 2002. Intelligent non-visual navigation of complex HTML structures. In Universal Access in the Information Society Volume 2, Number 1 (2002), Pages 56 – 69

128. Portigal, S., & Carey, T., (1994) Auralization of document structure. MSc. Thesis, The University of Guelph, Canada.

129. Raggett, D., Le Hors, A., Jacobs, I., 1999. HTML 4.01 Specification. http://www.w3.org/TR/html4, visited March 2010

130. Ram, S. and Sharf, J., "The People Sensor: A Mobility Aid for the Visually Impaired", in 2nd. International Symposium of Wearable Computers, IEEE (1998), pp 166-167.

131. Ramloll, R., Brewster, S.A., Yu, W., and Riedel, B., "Using non-speech sounds to improve access to 2D tabular numerical information for visually impaired users", presented at People and Computers XV - Interaction without Frontiers. Joint Proceedings of HCI 2001 and IHM 2001, 10 - 14 Sept. 2001, Lille, France, 2001.

132. Rayner, K. & Pollatsek, A. (1989). The Psychology of Reading. Englewood Cliffs, New Jersey: Prentice-Hall International, Inc.

133. Richards, J., Spellman, J., Treviranus, J., May, M., 2015. W3C WAI Authoring Tool Accessibility Guidelines (ATAG) 2.0. http://www.w3.org/TR/ATAG20/, visited November 2018

134. Roffler, S., and Butler, R., "Factors that influence the localization of sound in the vertical plane", Journal of the Acoustical Society of America, vol. 43, pp. 1260-1266, 1967.

135. Ross, D.A. and Blasch, B.B., "Wearable interfaces for orientation and wayfinding", in Proceedings of ASSETS 2000, ACM Press (2000), pp 193-200.

136. Roth, P., Petrucci, L., Assimacopoulos, A., Pun, T. “AB-Web: Active audio browser for visually impaired and blind users”. In International Conference on Auditory Display (ICAD'98). Glasgow, UK, November 1998.

137. Roth, P., Petrucci, L., Pun, T., Assimacopoulos, A. “Auditory browser for blind and visually impaired users”. In: ACM SIGCHI Computer Human Interaction 99, Conference on Human Factors in Computing Systems. 1999. p. 218-219

138. Rovers, A.F. and Van Essen, H.A., "HIM: A Framework for Haptic Instant Messaging", in Extended Abstracts of CHI 2004, ACM Press (2004), pp 1313-1316.

139. Rovers, A.F. and van Essen, H.A., "FootIO – Design and evaluation of a device to enable foot interaction over a computer network", in Proceedings of World Haptics 2005, IEEE Press (2005), pp 521-522.

140. Sakamoto, N., Gotoh, T. & Kimaura, Y. (1976). On 'out of head localization' in headphone listening. Journal of the Acoustic Engineering Society, 24(9), pp. 710 716.

141. Scharf, B. & Houtsma, A.J. (1986). Audition II: Loudness, pitch, localization, aural distortion and pathology. In K. R. Boff, L. Kaufman, & P. Thomas (Eds.), Handbook of perception and human performance, pp. 15.1-15.60. New York: Wiley.

142. Schick, A. & Meis, M.: Endbericht „Zugang zum Internet für Blinde und Sehgeschädigte - Unterauftrag AP3 und AP4. Interner Bericht des Projekts "Zugang zum Internet für Blinde", 2002.

143. Scholes, P.A. (1975). The oxford companion to music (10 th ed.). Oxford: Oxford University Press.

144. Sharmin, S., Evreinov, G., and Raisamo, R. Non-Visual Feedback Cues for Pen Comuputing. In Proc. Eurohaptics, IEEE Computer Society (2005), 625 - 628.

145. Sherrick, C. E. and Cholewiak, R. W., "Cutaneous Sensitivity," in Handbook of Perception and Human Performance. Eds. K. Boff, L. Kaufman, and J. L. Thomas. New York: Wiley and Sons, 1986, Ch 12 pp. 1 - 58.

146. Shneiderman, B. (1998): Designing the user interface, 3 rd Ed. Addison-Wesley, Reading (MA).

147. Slowiaczek, L.M. & Nusbaum, H.C. (1985). Effects of speech rate and pitch contour on the perception of synthetic speech. Human Factors, 27(6), pp. 701 712.

148. Smither, J. (1993). Short term memory demands in processing synthetic speech by old and young adults. Behaviour and Information Technology, 12(6), pp. 330-335.

149. Smyth, J. D., Dillman, D. A., Christian, L. M., & Stern, M. J. (2006). Comparing check-all and forced-choice question formats in web surveys. Public Opinion Quarterly, 70(1), 66-77.

150. Sonnenwald, D.H., Gopinath, B., Haberman, G.O., Keese, W.M. & Myers, J.S. (1990). InfoSound: An audio aid to program comprehension. Proceedings of the 23rd Hawaii International Conference on System Sciences, pp. 541-546.

151. Sorin, L., Lemarié, J., Aussenac-Gilles, N., Mojahid, M. & Oriola, B. (2014). Communicating text structure to blind people with Text-To-Speech. In: International Conference on Computers Helping People with Special Needs - ICCHP 2014, 9 July 2014 - 11 July 2014 (Paris, France).

152. Stevens, R.D., Brewster, S.A., Wright, P.C. & Edwards, A.D.N. (1994). Providing an audio glance at algebra for blind readers. In G. Kramer (Ed.), Accepted for publication in the proceedings of ICAD'94, Santa Fé Institute, Santa Fé: Addison-Wesley.

153. Strybel, T., Manligas, C. & Perrott, D. (1992). Minimum audible movement angle as a function of the azimuth and elevation of the source. Human Factors, 34(3), pp. 267-275.

154. Sumikawa, D., Blattner, M., Joy, K. & Greenberg, R. (1986). Guidelines for the syntactic design of audio cues in computer interfaces (Technical Report No. UCRL 92925). Lawrence Livermore National Laboratory.

155. Sumikawa, D.A. (1985). Guidelines for the integration of audio cues into computer user interfaces (Technical Report No. UCRL 53656). Lawrence Livermore National Laboratory.

156. Vanderveer, N.J. (1979). Ecological acoustics: Human perception of environmental sounds (Thesis No. 40/09B, 4543). Dissertation Abstracts International.

157. Van Erp, J.B.F. and Van Veen, H.A.H.C., "Vibro-Tactile Information Presentation in Automobiles", in Proceedings of Eurohaptics 2001, (2001), pp 99-104.

158. Van Erp, J.B.F., "Guidelines for the Use of Vibro-Tactile Displays in Human Computer Interaction", in Proceedings of Eurohaptics 2002, (2002), pp 18-22.

159. Van Erp, J.B.F. and Spapé, M.M.A., "Distilling the Underlying Dimensions of Tactile Melodies", in Proceedings of Eurohaptics 2003, (2003), pp 111-120.

160. Van Kesteren, A., Aryeh, G., Russell, A., Berjon, R. (2015). W3C Document Object Model (DOM) 4. http://www.w3.org/TR/dom/, visited November 2018

161. Van Veen, H.A.H.C. and van Erp, J.B.F., "Tactile Information Presentation in the Cockpit." in Proceedings of the First International Workshop on Haptic Human-Computer Interaction, Springer LNCS (2000), pp 174-181.

162. Vitense, H. S., Jacko, J. A., and Emery, V. K. Multimodal feedback: an assessment of performance and mental workload. Ergonomics 46(1) (2003), 68 - 87.

163. Wang, Q., Hayward, V.: Compact, Portable, Modular, High-performance, Distributed Tactile Transducer Device Based on Lateral Skin Deformation. Proceedings of the Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (HAPTICS'06), pp. 67-72. IEEE Computer Society, Arlington, VA, USA (2006)

164. Walker, B. N., Nance, A., & Lindsay, J. (2006). Spearcons: Speech-based earcons improve navigation performance in auditory menus. Proceedings of the 12th International Conference on Auditory Display, London, England.

165. Warren, W.H. & Verbrugge, R.R. (1984). Auditory perception of breaking and bouncing events: A case study in ecological acoustics. Journal of Experimental Psychology: Human Perception and Performance, 10, pp. 704-712.

166. WBU current national, special, and grouped member organizations http://www.worldblindunion.org/English/about-wbu/membership/Pages/Member-Regions-and-Organizations.aspx, visited April 2015

167. Wenzel, E.M., Foster, S.H., Wightman, F.L. & Kistler, D.J. (1989). Realtime digital synthesis of localized auditory cues over headphones. In IEEE workshop on applications of signal processing to audio & acoustics. Oct. 15-18, New Paltz, N.Y.

168. Wenzel, E., Wightman, F.L. & Kistler, D. (1991). Localization with non-individualized virtual display cues. In S. Robertson, G. Olson, & J. Olson (Eds.), CHI’91, New Orleans: ACM Press, Addison-Wesley, pp. 351-359.

169. WHO International Classification of Diseases ICD http://www.who.int/classifications/icd/en/, visited March 2010

170. Wickens, C.D., Mountford, S.J. & Schreiner, W. (1981). Multiple resources, task-hemispheric integrity and individual differences in time-sharing. Human Factors, 23(2), pp. 211-229.

171. Wood, L., Nicol, G., Robie, J., Champion, M., & Byrne, S. (2004). Document Object Model (DOM) level 3 core specification.

8.2 Products

- [A11y] Open Accessibility (A11y) Workgroup
http://www.a11y.org, visited July 2010
- [AppleZoom] Apple, Zoom for MacOS
http://www.apple.com/accessibility/macosx/vision.html, visited February 2010
- [AppleZoomIOS] Apple, Zoom for iOS
http://www.apple.com/iphone/how-to/index.html#accessibility.zoom, visited January 2010
- [Dragon] Nuance, Dragon NaturallySpeaking Speech Recognition Software
http://www.nuance.com/dragon/index.htm, visited January 2012
- [FineReader] ABBYY, FineReader OCR software for text recognition
http://finereader.abbyy.com, visited January 2012
- [JAWS] Freedom Scientific, JAWS for Windows Screen Reading Software
http://www.freedomscientific.com/products/fs/jaws-product-page.asp, visited September 2011
- [Jinput] Java.net, Jinput Java Game Controller API
http://java.net/projects/jinput, visited July 2010
- [Lunar] Dolphin, Lunar Magnifier
http://www.dolphinuk.co.uk/productdetail.asp?id=3, visited February 2010
- [Magic] Freedom Scientific, MAGic
http://www.freedomscientific.com/products/lv/magic-bl-product-page.asp, visited August 2011
- [MobileSpeak] CodeFactory, Mobile Speak
http://www.wireless.att.com/about/disability-resources/mobile-speak-magnifier.jsp, visited February 2011
- [NVDA] Nonvisual Desktop Access NVDA
http://www.nvda-project.org, visited January 2011
- [OmniPage] Nuance, OmniPage sacnning and OCR software
http://www.nuance.com/for-individuals/by-product/omnipage, visited January 2012
- [OpenBook] Freedom Scientific, OpenBook Scanning and Reading Software
http://www.freedomscientific.com/products/fs/openbook-product-page.asp, visited February 2011
- [Rumblepad] Logitech, Rumblepad 2 Vibration Feedback PC-Gamepad
http://www.logitech.com/en-us/support/game-gear/264, visited June 2011
- [Series7Slate] Samsung, Series 7 Slate PC
http://www.samsung.com/uk/consumer/pc-peripherals/notebook-computers/ultra-portable/XE700T1A-A01UK, visited January 3023
- [Talks] Nuance, Talks
http://www.nuance.com/talks, visited March 2011
- [ViewPlus] Braille Printers (Braille Embossers) and Assistive Technology Products
http://www.viewplus.com, visited April 2011

- [VoiceOver] Apple, VoiceOver
  http://www.apple.com/accessibility/voiceover, visited January 2010
- [VoiceOverIOS] Apple, VoiceOver for iOS
  http://www.apple.com/accessibility/voiceover, visited January 2012
- [VoicePro] Linguatec, Voice Pro
  http://www.linguatec.net/products/stt/voice_pro, visited January 2012
- [WindowsMagnifier] Microsoft, Windows Magnifier
  http://www.microsoft.com/enable/training/windowsvista/magnifier.aspx, visited Fabruary 2010
- [WinMobileLensPro] Adisasda, WinMobile Lens Pro
  http://www.adisasta.com/wmLensPro.html, visited March 2011
- [XYaxesPDF] xyMedia, axesPDF for Word
  http://www.axespdf.com, visited June 2012
- [ZHAWAccessiblePDF] ZHAW InIT, Plug-ins for accessible PDF documents
  http://www.init.zhaw.ch/en/engineering/institute-of-applied-information-technology/focus-areas/human-information-interaction/plug-ins-for-accessible-pdf-documents.html, visited March 2012
- [Zooms] Nuance, Zooms
  http://www.nuance.com/zooms, visited February 2011
- [ZoomText] Aisquared, ZoomText
  http://www.aisquared.com/zoomtext, visited Fabruary 2010
- [ZoomTextReader] Asquared, ZoomText Reader
  http://www.aisquared.com/zoomtext, visited March 2011

Appendix A: Survey Questionnaire

The following questionnaire has been used for the initial online internet survey described in Chapter 4:

Survey: How do blind and visually impaired people handle text-documents ?

Introduction

Dear Sir or Madam

My name is Martin Dorigo. After successfully completed my studies in computer science I'm now intending to write a dissertation in the field of accessibility. The goal of my project is to develope a novel concept for an electronic assistative technology in support of blind and visually impaired people making text-documents more accessible to them. Self concerned, I'm faced with the big challenge of reading text-documents every day.

With this questionaire I would like to find out how you handle text-documents, which issues you have and how you sort these problems out at the moment. In addition I am interested what is of importance for you as to the reading of text and the navigation within documents and which features you wish for a novel assistative technology. For me text-documents are: Letters, newspapers, journals, books, teaching material, websites or right this questionaire.

To provide a basis for the research und development of an adjuvant instrument for the benefit of blind and visually impaired, I'm reliant on your large und significant feedback. Please do not hesitate to give a detailed description on the comment field if your answer is not covered by the choices available. Please take a few minutes to answer my questionaire:

Researcher

Martin Dorigo
University of Plymouth
Drake Circus
Plymouth
PL4 8AA
Email: **martin.dorigo@plymouth.ac.uk**

Supervised by

- Prof. Dr. Bettina Harriehausen-Mühlbauer
  University of Applied Sciences Darmstadt
  Email: **b.harriehausen@fbi.h-da.de**
- Dr. Ingo Stengel
  University of Applied Sciences Darmstadt
  Email: **ingo.stengel@h-da.de**
- Dr. Paul Dowland
  University of Plymouth
  Email: **p.dowland@plymouth.ac.uk**

It is up to you to decide whether or not to take part and you are free to withdraw at any time, without giving reason. All information collected for this study will be kept strictly confidential and full anonymity of participants will be ensured during the collection, storage and publication of research material. Participants can request a copy of the published research. If you are not 18 years or older, please do not answer this survey.

☐ I confirm that I have read and understand the information for the above study and agree to take part and that I'm 18 years old and above. I agree to the use of anonymised quotes in publications.

Questionaire

1. What do you do with documents ?

- Read documents
  Never ◯ Seldom ◯ Often ◯ Very often ◯
- Compose new documents
  Never ◯ Seldom ◯ Often ◯ Very often ◯
- Edit existing documents
  Never ◯ Seldom ◯ Often ◯ Very often ◯

Comment / What else do you do with documents ?

2. In which form are the document you would like to read provided to you ?

- Printed paper
  Never Seldom Often Very often
- Braille paper
  Never Seldom Often Very often
- Electronically
  Never Seldom Often Very often
- Acoustically (e.g. audiobook, podcast)
  Never Seldom Often Very often

If electronically: Which electronic formats ?

- PDF
  Never Seldom Often Very often
- Websites, E-Mail (HTML)
  Never Seldom Often Very often
- Microsoft Word
  Never Seldom Often Very often
- Microsoft Excel
  Never Seldom Often Very often
- Microsoft PowerPoint
  Never Seldom Often Very often
- RTF
  Never Seldom Often Very often
- Plain text
  Never Seldom Often Very often

Comment / In which other forms and formats are documents provided to you ?

3. In which form do you wish the documents you would like to read ?

- Printed paper
- Braille paper
- Electronically
- Acoustically
- Other

If electronically: Which electronic format ?

- PDF
- HTML
- Microsoft Word
- Microsoft Excel
- Microsoft PowerPoint
- RTF
- Plain text
- Other

Comment / In which other forms and formats do you wish the documents you would like to read ?

4. Which assistative technologies do you use ?

For paper based documents

- Magnifying glass
  Yes ◯ No ◯
- Electronic magnifier
  Yes ◯ No ◯
- Electronic reader
  Yes ◯ No ◯

For computer

- Screen magnifier
  Yes ◯ No ◯
- Screen reader
  Yes ◯ No ◯
- Braille display
  Yes ◯ No ◯

For mobile phone

- Screen magnifier
  Yes ◯ No ◯
- Screen reader
  Yes ◯ No ◯

Comment / Which other assistative technologies do you use ?

5. Which input devices do you use ?

For computer

- Mouse
  Yes No
- Keyboard
  Yes No
- Joystick
  Yes No
- Speech recognition
  Yes No
- Touchscreen
  Yes No
- Gestures
  Yes No

For mobile phone

- Keyboard
  Yes No
- Joystick
  Yes No
- Speech recognition
  Yes No
- Touchscreen
  Yes No
- Gestures
  Yes No

Comment / Which other input devices do you use ?

6. How important are the following structural elements for you to navigate within a document ?

- Headings
  Unimportant Less important Important Very important
- Bookmarks
  Unimportant Less important Important Very important
- Paragraphs
  Unimportant Less important Important Very important
- Lists
  Unimportant Less important Important Very important
- Tables
  Unimportant Less important Important Very important
- Form fields
  Unimportant Less important Important Very important
- Graphics
  Unimportant Less important Important Very important
- Links
  Unimportant Less important Important Very important

Comment / Which other structural informations do you use to navigate within a document ?

7. How satisfied are you with your current situation for reading documents ?

- Unsatisfied: The reading of documents is very difficult for me. I always fail in finding the relevante information in time. I feel poor informed.
- Adequate: I can read the absolutely necessary information. I feel adequate informed.
- Very satisfied: I can read documents quick and easy and locate the relevant information fast. I feel informed very well.

Comment

8. If you encounter any difficulties during the reading of a document, how do you solve that problem at the moment ?

- Somebody reads the document to me
- A normal sighted person makes the document accessible for me manually
- A software or device makes the document accessible automatically
- I can make the document accessible myself (with great effort)
- Unfortunately I do not have a solution at the moment

Comment / Which other solutions do you have ?

9. Which problems do you have as to the reading of documents and how often do they occur ?

- Wrong reading order
  Never Seldom Often Very often
- Document language not or wrong defined
  Never Seldom Often Very often
- I can not make the characters out (e.g. font or special characters)
  Never Seldom Often Very often
- Contrast ratio between text and background insufficient
  Never Seldom Often Very often
- Accentuations are not recognisable for me
  Never Seldom Often Very often
- Document does not provide structural information
  Never Seldom Often Very often
- Layout-tables (Tables are used for layout and not for semantic data belonging together)
  Never Seldom Often Very often
- Graphics are not recognisable and do not have an alternative text
  Never Seldom Often Very often
- Copy protection avoids access to the document for assistative technologies
  Never Seldom Often Very often

Comment / Which other problems do you have as to reading documents and how often do they occur ?

10. Which features do you wish for a novel assistative technology ?

Please give a detailed description of your wishes and needs.

11. Which electronic devices do you have at the moment and which one(s) do you intend to purchase within 2 to 3 years ?

At the moment

- Desktop computer
  Yes No
- Portable computer
  Yes No
- Mobile phone
  Yes No
- Mobile phone with camera
  Yes No
- Mobile phone with internet access
  Yes No
- eBook reader
  Yes No
- Daisy player
  Yes No
- Scanner
  Yes No

Within 2 to 3 years

- Desktop computer
  Yes No
- Portable computer
  Yes No
- Mobile phone
  Yes No
- Mobile phone with camera
  Yes No
- Mobile phone with internet access
  Yes No
- eBook reader
  Yes No
- Daisy player
  Yes No
- Scanner
  Yes No

Comment / Which other devices do you have already and which one(s) are you going to purchase ?

12. From which visual impairments do you suffer ?

- Visual acuity -- Please select --
- Restricted visual field
  Yes No
- Visually impaired by birth
  Yes No

Please give a detailed description of your visual impairment.

13. About your person

Occupational situation -- Please select -- Occupation

Highest education -- Please select --
Computer literacy -- Please select --
Gender -- Please select --
Age

The following information is optional. It is needed for further enquiry if necessary and will not be passed to third parties. This questionaire can also be answered anonymous.

First name Family name
Address
Phone
E-Mail
Comment

14. Would you like to participate in a pilot test ?

- Yes please. I would like to make a contribution to the problem solving and I can benefit from noval research findings at first hand.
- No thanks. I'm not interested in a cooperation.

Comment

Thank for completing my questions. With your valuable assistance you made a big contribution to the research in support of blind and visually impaired people.

Submit questionaire

Appendix B: Observation Schedule

Briefing

Research Goal

Instructions:

Hello. My name is Doky.

In the next 15 Minutes I would like to let you try your hand at a novel system for reading and editing structured documents in a nonvisual way on mobile and wearable devices.

It enables you to get a fast overview over the document structure and to efficiently skim and scan over the document content as well as to move, remove and insert elements and text.

Continuation:

Please tap with one finger anywhere on the screen to continue.

Procedure

Instructions:

After this introduction, you will perform 19 exercises organised in 3 parts.
Each exercise consists of instructions and a task you have to perform among an example document.
You can listen to the instructions again at any time by double-tapping with one finger anywhere on the screen.
While you are solving a task the screen is black and I will record your actions performed and time taken.
Finally, you will be debriefed with your results.

Continuation:

Please tap with one finger anywhere on the screen to continue.

Informed Consent

Instructions:

It is up to you to decide whether or not to take part and you are free to withdraw

at any time without giving reason by pressing the home button.

All information collected for this study will be kept strictly confidential and full anonymity of participants will be ensured during the collection, storage and publication of research material.

Participants can request a copy of the published research. If you are not 18 years or older, please do not take part.

When you take part you confirm that you have understand the information for this research and agree to take part and that you are 18 years old and above.

Continuation:
Please tap with one finger anywhere on the screen to continue.

Part 1: Navigation

Instructions:
In the first 7 exercises you will learn how to navigate within a document.
There are two different ways for doing this:
You can either use your fingers on the screen.
Or you can move the device itself.
In the instructions always both ways are explained to you.
Afterwards it is up to you to decide which way to use to solve an exercise.

Continuation:
Please tap with one finger anywhere on the screen to start with the first exercise.

Set Element Cursor

Instructions:
To go into the document:
Put one finger on the screen.
Or rotate the device horizontally anticlockwise 10 degrees.

Task:
Please go into the document.

Congratulation:
Congratulations, you have successfully gone into the document.

Continuation:
Please tap with one finger anywhere on the screen to continue with the next exercise.

Move Element Cursor

Instructions:
When you are inside the document, you can move forward or backward at an arbitrary speed:
Slide the finger up or down.
Or rotate the device horizontally clockwise or anticlockwise within an angle of 180 degrees.

Task:
Please move from the beginning to the end of the document.

Congratulation:
Congratulations, you have successfully solved this exercise.

Continuation:
Please tap with one finger anywhere on the screen to continue with the next exercise.

Set Text Cursor

Instructions:
When you are at an element containing text, you can go into this text:
Move the finger 1 centimeter to the left or to the right.
Or rotate the device vertically anticlockwise 10 degrees.

Task:
Please go into an arbitrary text.

Congratulation:
Congratulations, you have successfully gone into a text.

Continuation:
Please tap with one finger anywhere on the screen to continue with the next exercise.

Move Text Cursor

Instructions:
When you are inside a text, you can move by word forward and backward at an arbitrary speed:
Slide the finger left or right.
Or rotate the device vertically clockwise or anticlockwise within an angle of 90 degrees.

Task:
Please move from the beginning to the end of an arbitrary text.

Congratulation:
Congratulations, you have successfully solved this exercise.

Continuation:
Please tap with one finger anywhere on the screen to continue with the next exercise.

Unset Text Cursor

Instructions:
When you are inside a text, you can go out of this text:
Move the finger more than 1 centimeter up or down.
Or rotate the device vertically back to the starting position.

Task:
Please go out of an arbitrary text.

Congratulation:
Congratulations, you have successfully gone out of a text.

Continuation:

Please tap with one finger anywhere on the screen to continue with the next exercise.

Unset Element Cursor

Instructions:

When you are inside the document, you can go out of the document:

Lift up the finger.

Or rotate the device horizontally back to the starting position.

Task:

Please go out of the document.

Congratulation:

Congratulations, you have successfully gone out of the document.

Continuation:

Please tap with one finger anywhere on the screen to continue with the next exercise.

Set Select Modifier

Instructions:

You can select the element or text at your current position:

Put down a second finger anywhere on the screen.

Or flip the device 90 degrees to the left.

Task:

Please select an arbitrary element or text.

Congratulation:

Congratulations, you have successfully solved this exercise.

Continuation:

Please tap with one finger anywhere on the screen to continue with the next part.

Part 2: Presentation

Instructions:

In the next 6 exercises you will learn how the document is presented to you.

There are three different ways in which this is done:

You hear tones.

You hear speech.

And you feel vibrations.

In the instructions always all ways are explained to you.

Afterwards it is up to you to decide which ways to use to solve an exercise.

Continuation:
Please tap with one finger anywhere on the screen to start with the first exercise.

Element Type

Instructions:
A heading is acoustically presented by trumpet, a section by brass, a paragraph by guitar, a graphic by glockenspiel, a list by horn, a list-item by oboe, a table by organ, a table row by violin and a table cell by clarinet.

Using vibrations, a heading is presented by one eighth and one full note, a section by 3 eighth notes, a paragraph by one quarter and one eighth note, a graphic by one full one eighth and one half note, an list by one full and one half note, a list-item by one quarter one half and one eighth note, a table by one quarter one half and one full note, a table row by one quarter one eighth and one quarter note, and a table cell by one quarter one eighth and one full note.

Task:
Please select a graphic.

Congratulation:
Congratulations, you have successfully solved this exercise.

Continuation:
Please tap with one finger anywhere on the screen to continue with the next exercise.

Hierarchy Level

Instructions:
For each hierarchy level, the pitch of the tone and the speech and the tempo of the vibrations is increased.

Task:
Please select an element at the 4th level.

Congratulation:
Congratulations, you have successfully solved this exercise.

Continuation:
Please tap with one finger anywhere on the screen to continue with the next exercise.

Position

Instructions:
The relative position of an element or text on the screen and on the clock represents the position of it within the document.

Task:
Please select the element which is in the middle of the document.

Congratulation:
Congratulations, you have successfully solved this exercise.

Continuation:
Please tap with one finger anywhere on the screen to continue with the next exercise.

Length

Instructions:
The relative length of an element on the screen or on the clock represents the length of the text contained in it.

Task:
Please select the element containing the longest text.

Congratulation:
Congratulations, you have successfully solved this exercise.

Continuation:
Please tap with one finger anywhere on the screen to continue with the next exercise.

Relationship

Instructions:
You can hear the tones for all nested structure elements and text at your current position simultaneously.

And you can hear the speech and feel the vibrations of them sequentially.

Task:
Please select the paragraph which is inside 3 sections.

Congratulation:
Congratulations, you have successfully solved this exercise.

Continuation:
Please tap with one finger anywhere on the screen to continue with the next exercise.

Element Text

Instructions:
The text contained in an element is spoken. If an element has no text but contains a heading, the text of this heading is spoken instead.

Task:
Please select the element which contains the text "Strawberry".

Congratulation:
Congratulations, you have successfully solved this exercise.

Continuation:

Please tap with one finger anywhere on the screen to continue with the next part.

Part 3: Manipulation

Instructions:

In the last 6 exercises you will learn how to activate, move, remove and insert structure elements and text in a document.

Continuation:

Please tap with one finger anywhere on the screen to start with the first exercise.

Activate Selection

Instructions:

When a selection is selected, you can activate it:

Put down a third finger anywhere on the screen.

Or flip the device 180 degrees to the left.

The system presents this selection acoustically at a lower register and using vibrations with a lower intensity.

Task:

Please activate an arbitrary selection.

Congratulation:

Congratulations, you have successfully activated a selection.

Continuation:

Please tap with one finger anywhere on the screen to continue with the next exercise.

Deactivate Selection

Instructions:

When a selection is active, you can deactivate it:

Lift up this third finger.

Or flip back the device 90 degrees to the right.

The system stops presenting the selection at a lower register and with a lower intensity.

Task:

Please deactivate an arbitrary selection.

Congratulation:

Congratulations, you have successfully deactivated a selection.

Continuation:

Please tap with one finger anywhere on the screen to continue with the next exercise.

Move Selection

Instructions:

When a selection is active, you can move it:

Move to the target position.

The system presents the selection acoustically at a lower register and using vibrations with a lower intensity.

Deactivate the selection afterwards.

Task:

Please move an arbitrary selection to another position.

Congratulation:

Congratulations, you have successfully moved a selection.

Continuation:

Please tap with one finger anywhere on the screen to continue with the next exercise.

Remove Selection

Instructions:

When a selection is active, you can remove it:

Go out of the document.

The system presents the selection acoustically at a lower register and using vibrations with a lower intensity.

Deactivate the selection afterwards.

Task:

Please remove an arbitrary selection.

Congratulation:

Congratulations, you have successfully removed a selection.

Continuation:

Please tap with one finger anywhere on the screen to continue with the next exercise.

Insert Element

Instructions:

You can insert a new structure element at your current position:

Slide in a second finger from the bottom into the screen or flip the device 90 degrees to the right.

Slide this second finger up or down or rotate the device horizontally clockwise or anticlockwise within an angle of 180 degrees to choose the element type you want to insert.

The system presents the new element acoustically at a higher register and using vibrations with a higher intensity.

Select this element afterwards.

Task:
Please insert an arbitrary element into the document.

Congratulation:
Congratulations, you have successfully inserted a new element.

Continuation:
Please tap with one finger anywhere on the screen to continue with the last exercise.

Insert Text

Instructions:
You can insert new text at your current position:
Slide in a second finger from the left into the screen or flip the device 90 degrees to the right and rotate it 10 degrees vertically anticlockwise.
Speak the text you want to insert.
Afterwards lift up this second finger or flip back the device 90 degrees to the left.

Task:
Please insert an arbitrary text into the document.

Congratulation:
Congratulations, you have successfully inserted a text into the document.

Continuation:
This was the last Exercise. Please tap with one finger anywhere on the screen to continue with the debriefing.

Debriefing

Instructions:
Well done! You have successfully completed all exercises.
Thank you for taking part in this research. With your valuable assistance you made a big contribution to the research in support of blind and visually impaired persons.
The published results will be made available for download over the Internet or will be sent out to the participants by E-Mail.
For me it is now time to say good bye.

Continuation:
Please tap with one finger anywhere on the screen to terminate this app.

DOKY iPhone App: Non-Visual Reading and Editing Structured Documents

Dear Sir or Madam

In the frame of my Ph.D. Project I developed a novel system for reading and editing structured documents in a non-visual way on mobile and wearable devices. This App is called DOKY and it is available for Apple iPhone, iPad and Android devices!

I would like to invite as many of your members as possible to test my system. May I ask you to support me in the following two points:

1. Invite your members to test my app DOKY. This can be done by E-Mail list, newsletter, forum or on your website. I attached an invitation message below.
2. Inform other Organisations and forward this message to them.

Thank you so much for your valuable assistance.

Kind regards

Martin Dorigo

DOKY iPhone App: Non-Visual Reading and Editing Structured Documents

Dear Member

In the frame of his Ph.D. Project Martin Dorigo developed a novel system for reading and editing structured documents in a non-visual way on mobile and wearable devices. This App is called DOKY and it is available for Apple iPhone, iPad and Android devices!

DOKY provides you amongst others the following 3 unique key features:

1. A fast overview over the document structure
2. A fast skim and scan over the document content
3. Move, remove and insert elements and text

He would like to invite you to test his system. It will take you 15 minutes only:

1. Please go to the Apple AppStore or Android PlayStore and search for DOKY (written with Y).
   Or use the following links:
   http://itunes.apple.com/gb/app/doky/id975186203
   http://play.google.com/store/apps/details?id=ch.dorigo.doky
2. Install and start the DOKY app
3. Follow the spoken instructions

In order to reach significant results the participation of a lot of people is required. May I ask you to forward this message to many others.

Please test DOKY until April 30th 2015. With your valuable assistance the reading and editing of structured documents will be revolutionized!

Thank you so much and kind regards.

Appendix C: A Flexible Platform-Independent Event-Driven Software Architecture of the DOKY User Interface and Automated Structured Observation

Introduction

This chapter presents a flexible platform-independent and event-driven software architecture implementing the DOKY user interface proposed in the previous Chapter 5 as well as the automated structured observation research method employed for the evaluation of the effectiveness of the proposed user interface as described in detail in Chapter 3. Because it is platform- and language-neutral, it can be used in a wide variety of platforms, environments and applications for mobile and wearable devices like the Apple iOS, Google Android or Microsoft Windows Phone operating system.

Event-driven programming is a software paradigm in which the flow of the program is determined by events such as user actions, sensor outputs, or messages from other components of the program. Event-driven programming is the dominant paradigm employed in user interfaces and other applications that are centred on performing certain actions in response to user input.

In software architecture, publish-subscribe is an event pattern where senders of events, called publishers, do not program the events to be sent directly to specific receivers, called subscribers, but instead categorize published events into classes without knowledge of which subscribers, if any, there may be. Similarly, subscribers express interest in one or more classes and only receive events that are of interest, without knowledge of which publishers, if any, there are.

The software architecture is designed with flexibility in mind, so that it can easily by changed or extended, for example by adding alternative document models, input modalities, output modalities, observers, recorders, participants and observation schedules. Therefore, each component is defined by interfaces and abstract classes only and grouped in a semantically self-containing package. Figure C.1 gives an overview over the different components of this software architecture.

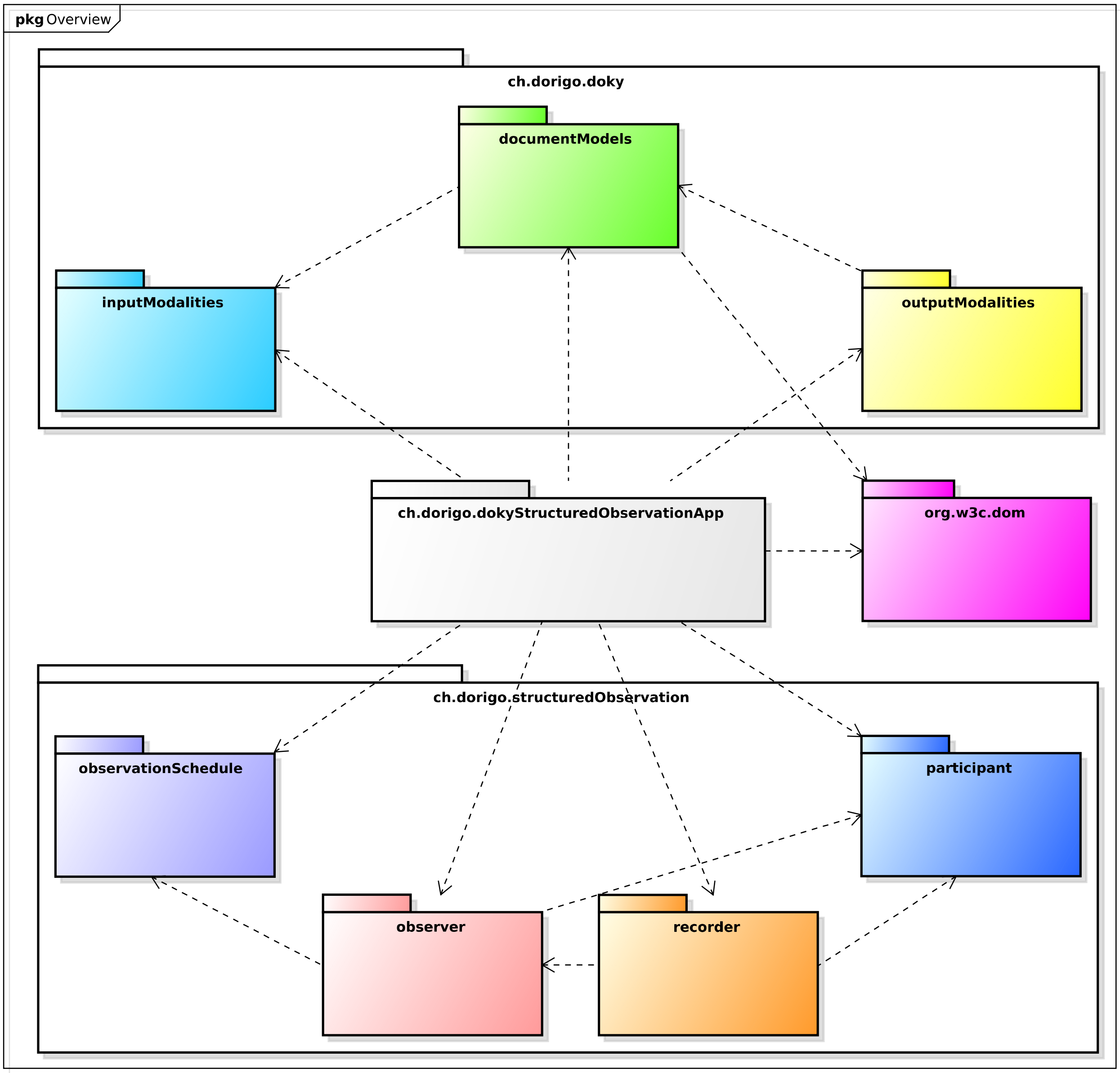


Figure C.1: Platform-Independent Event-Driven Software Architecture Overview

The Doky package implements the DOKY user interface, which is a multi-model user interface for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices. It contains sub-packages for all document models, input modalities and output modalities provided by the DOKY user interface.

The Document Models package contains all interfaces and classes of modality-neutral document models for presentation, navigation and manipulation of structured documents. A document model is a collection of descriptions of data structures and their contained fields, together with the operations or functions that manipulate them.

The Input Modalities package contains all interfaces and classes of input modalities. An input modality enables the user to perform the actions provided by the document models by stimulating s specific hardware human input device (HID).

The Output Modalities package contains all interfaces and classes of output modalities. An output modality presents the interactions of the document models to the user in a modality specific way by stimulation of one or more human senses using specific hardware human output devices.

The Structured Observation package implements the automated structured observation research method. Structured observation is a technique in which explicitly formulated rules for the observation and recording of behaviour are employed. The rules inform observers about what they should look for and how they should record behaviour. Each person who is part of the research is observed for a predetermined period of time using the same rules. These rules are articulated in what is usually referred to as an observation schedule. It contains sub-packages for all observers, recorders, participants and observation schedules provided by the automated structured observation research method.

The Observer package contains all interfaces and classes of observers. An observer applies the observation schedule to a research subject. It presents the title, instructions and task of an exercise to the research subject. Afterwards it observes the research subject while he or she is solving the task. If the task has been successfully solved, the congratulations and the continuation is presented.

The Recorder package contains all interfaces and classes of recorders. The aim of a recorder is to receive and record all events that occur in the user interface as well as in the automated structured observation process during the observation period, serialise an storing or streaming them over the Internet for later processing and analysis of this recorded data.

The Participant package contains all interfaces and classes of participants. A participant or research subject is a person who is taking part in the research and is observed for a predetermined period of time using the rules articulated in the observation schedule.

The Observation Schedule package contains all classes of an observation schedule. The aim of the observation schedule is to ensure that each participant’s behaviour is systematically recorded so that it is possible to aggregate the behaviour of all those in the sample in respect of each type of behaviour being recorded. The rules that constitute the observation schedule should be as specific as possible in order to direct observers to exactly what aspects of behaviour they are supposed to be looking for. Each person who is part of the research is observed for a predetermined period of time using the same rules.

The Doky Structured Observation App package contains the main class of the Doky Structured Observation App. The Doky Structured Observation App is an application of all the different components for conducting automated structured observations of research subjects performing exercises among a highly structured example document on their own mobile or wearable devices remotely over the Internet.

The Document Object Model package contains all interfaces of the Document Object Model (DOM). The Document Object Model (DOM) is a platform- and language-neutral application programming interface (API) that allows programs and scripts to dynamically access and update the content, structure and style of structured documents.

DOKY User Interface

The Doky package implements the DOKY user interface, which is a multi-model user interface for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices. The DOKY user interface has been described in detail in the previous Chapter 5. It contains sub-packages for all document models, input modalities and output modalities provided by the DOKY user interface.

Document Models

The Document Models package contains all interfaces and classes of modality-neutral document models for presentation, navigation and manipulation of structured documents. According to Van Kesteren, Aryeh, Russell & Berjon (2015), a document model is a collection of descriptions of data structures and their contained fields, together with the operations or functions that manipulate them. Figure C.2 gives an overview over the Document Models package:

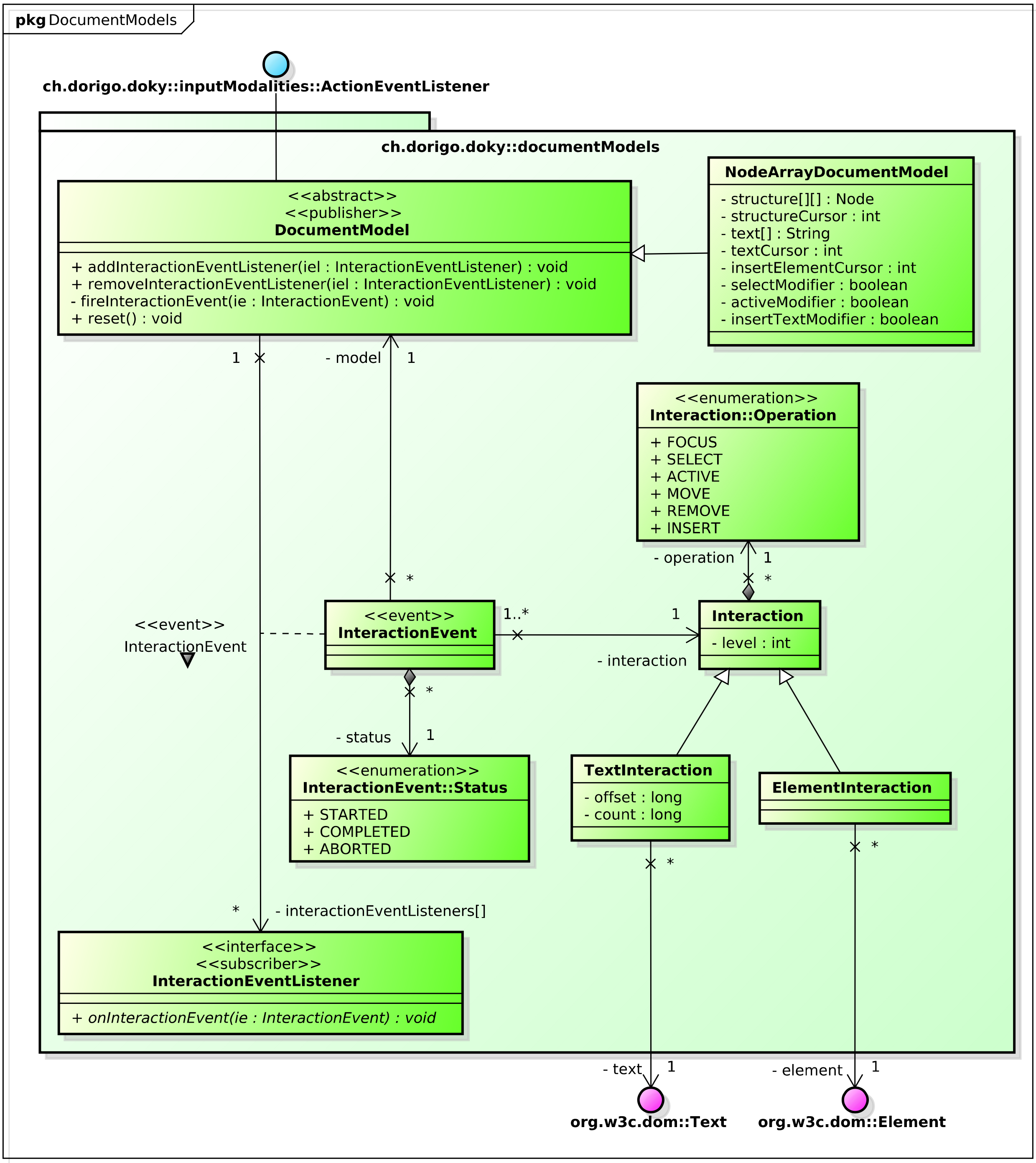


Figure C.2: Package Document Models Overview

The DocumentModel is the abstract base class for every document model. It implements the ActionEventListener interface of the InputModalities package, so that it can receive and react on ActionEvents performed by the user by employing different input modalities.

It can publish InteractionEvents and contains methods for adding and removing InteractionEventListeners as well as for firing interaction events and resetting the document model. It contains a collection of all registered interaction event listeners to which interaction events can be fired.

The InteractionEventListener is the listener interface for subscribers who want to receive interaction events of a document model. The class that is interested in processing interaction events implements this interface, and the object created with that class is registered with a document model, using it’s addInteractionEventListener method. When an interaction event occurs, that object's onInteractionEvent method is invoked.

The InteractionEvent is the event class which indicates that an interaction event occurred. It is generated by a document model when an interaction event occurs and passed to every interaction event listener that subscribed to receive interaction events. It contains a reference to the corresponding document model which is the source and has fired this interaction event. It also references the corresponding interaction on which the interaction event is performed. The interaction event also contains a status which is an enumeration and indicates whether the corresponding interaction has been started, completed or aborted.

The Interaction is the abstract base class for all interactions provided by document models. An interaction means performing an operation on a node of the document. It contains the operation which is an enumeration and can take on the values focus, select, active, move, remove and insert as well as the level of the node within the document, on which the operation of this interaction is performed. There are two different subclasses of interactions: ElementInteractions and TextInteractions.

The ElementInteraction is a subclass which extends the abstract Interaction base superclass. This is an interaction that is performed on an element node of the document. It contains a reference to the corresponding element node within the document, on which the operation of this interaction is performed.

The TextInteraction is an alternative subclass which also extends the abstract Interaction base superclass. This, on the other hand, is an interaction that is performed on a text node of the document. It contains a reference to the corresponding text node within the document, on which the operation of this interaction is performed. In addition, it contains an offset defining the start position as well as a count defining the number of characters to the end position within the text content of the text node involved.

The NodeArrayDocumentModel is a subclass which extends the abstract DocumentModel base superclass. It implements the Node Array Document Model, which is an alternative modality-neutral document model for presentation, navigation and manipulation of a structured documents. In contrast to the tree-like document representation employed in the Document Object Model (DOM), the element and text nodes of the document are organised in a two-dimensional array. The Node Array Document Model has been introduced in detail in Chapter 5.2.1.

Input Modalities

The Input Modalities package contains all interfaces and classes of input modalities. An input modality enables the user to perform the actions provided by the document models by stimulating s specific hardware human input device (HID). Figure C.3 gives an overview over the Input Modalities package:

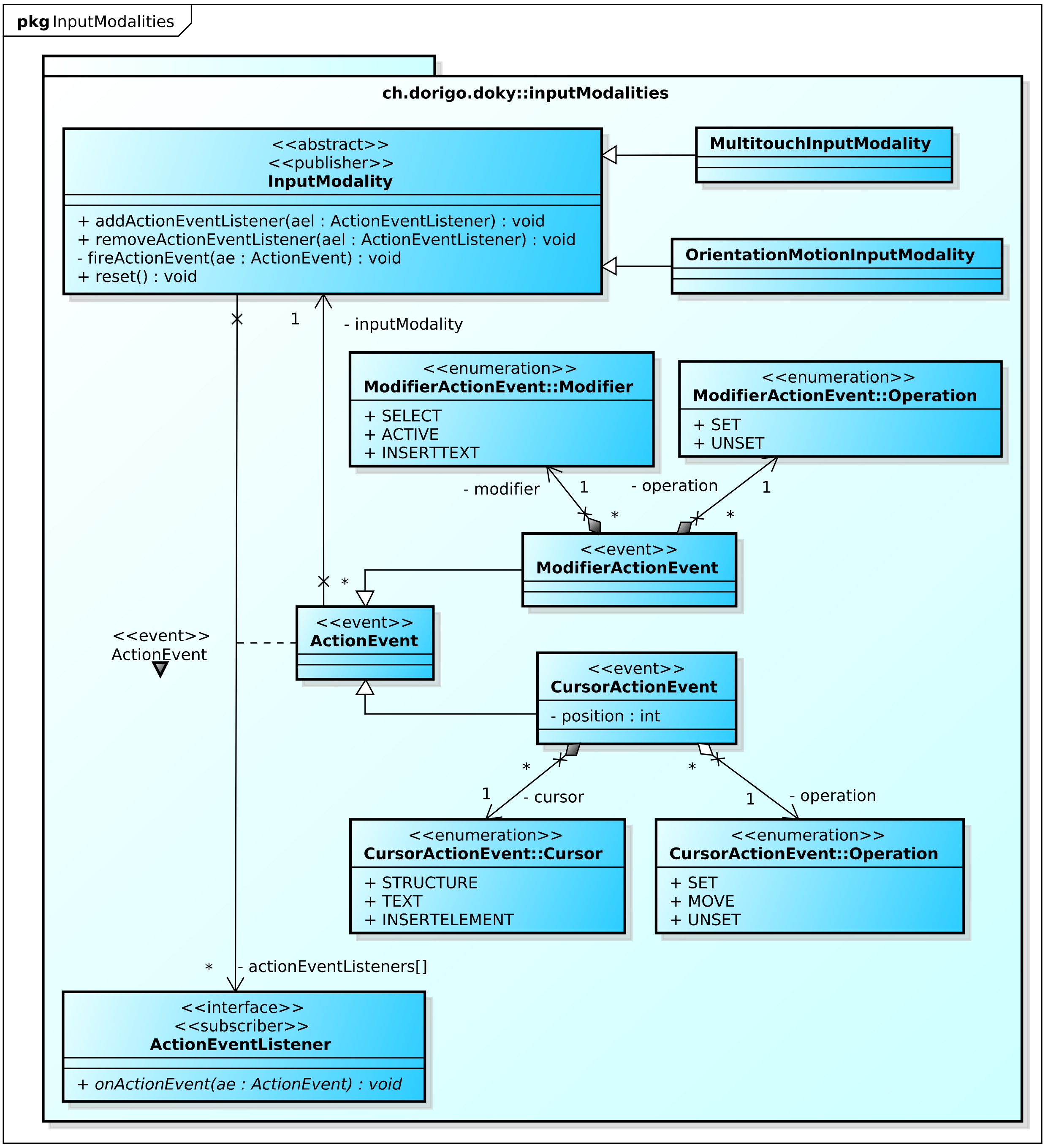


Figure C.3: Package Input Modalities Overview

The InputModality is the abstract base class for all input modalities. It can publish ActionEvents and contains methods for adding and removing ActionEventListeners as well as for firing action events and resetting the input modality. It contains a collection of all registered action event listeners to which action events can be fired.

The ActionEventListener is the listener interface for subscribers who want to receive action events of an input modality. The class that is interested in processing action events implements this interface, and the object created with that class is registered with an input modality, using it’s addActionEventListener method. When an action event occurs, that object's onActionEvent method is invoked.

The ActionEvent is the abstract base event class for all action events which indicates that an action has been performed by the user employing an input modality. It contains a reference to the corresponding input modality which is the source and has fired this action event. There are two subclasses of action events: CursorActionEvents and ModifierActionEvents.

A CursorActionEvent is an action event that is performed on an cursor of the document model. It contains the corresponding cursor which is an enumeration and can take on the values structure, text and insertElement. It also contains the position of the cursor on the corresponding axis in the document model. In addition, it also contains an operation which is an enumeration too and can take on the values set, move and unset.

A ModifierActionEvent, on the other hand, is an action event that is performed on a modifier of the document model. It contains the modifier which is an enumeration and can take on the values select, active and insertText. It also contains an operation which is an enumeration too and can take on the values set and unset.

The MultitouchInputModality is a subclass which extends the abstract InputModality base superclass. It implements the Multitouch Input Modality, which allows the user to perform the actions provided by the document models by performing multi-touch gestures on the multi-touch screen or multi-touch pad of the mobile and wearable device. With multi-touch input pointers act relative to coordinates of the device. This input modality is mainly purposed for smart-phones, smart-pablets and smart-tablets because these category of devices typically have a large multi-touch screen embedded and may be too large and heavy to be moved around themselves for using motion. The Multitouch Input Modality has been described in detail in Chapter 5.2.5.

The OrientationMotionInputModality is an alternative subclass which extends the abstract InputModality base superclass. It implements the Orientation Motion Input Modality, which allows the user to perform the actions provided by the document models by performing motion gestures by moving the mobile or wearable device itself. With motion input, the device itself is acting relative to coordinates of the word. The orientation of a device is typically calculated by using a combination of the linear acceleration and magnetic field sensor, what is known as sensor fusion. This input modality is mainly purposed for wearable devices like smart-watches because these category of devices typically are very small and lightweight and can therefore be moved around easily. In addition there may be only a very small multi-touch screen or multi-touch pad or no multi-touch screen or multi-touch pad at all embedded in these devices making the use of multi-touch input difficult. The Orientation Motion Input Modality has been described in detail in Chapter 5.2.6.

Output Modalities

The Output Modalities package contains all interfaces and classes of output modalities. An output modality presents the interactions of the document models to the user in a modality specific way by stimulation of one or more human senses using specific hardware human output devices. Figure C.4 gives an overview over the Output Modalities package:

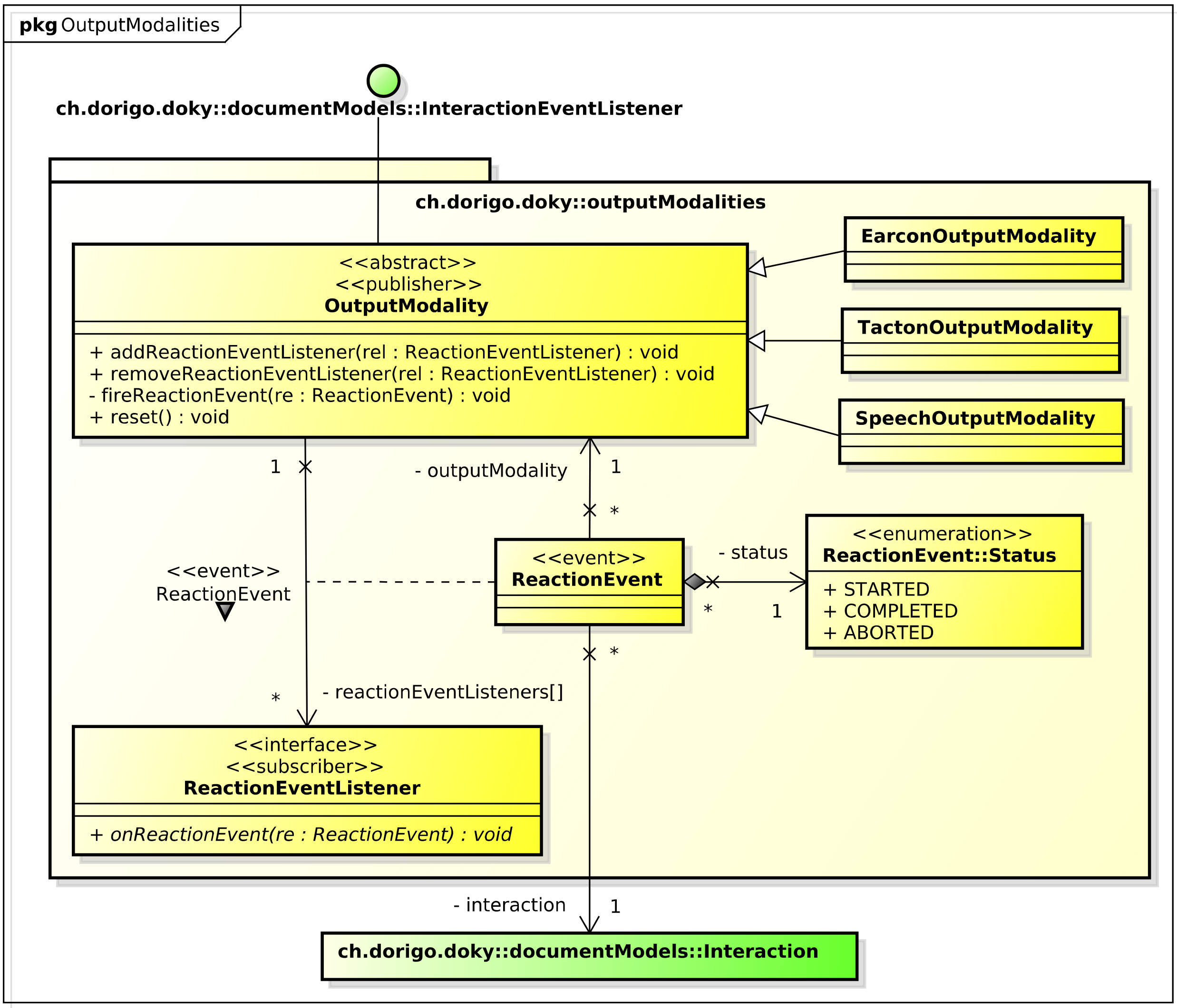


Figure C.4: Package Output Modalities Overview

The OutputModality is the abstract base class for all output modalities. It implements the InteractionEventListener interface of the Document Models package, so that it can receive and react on InteractionEvents provided by the document models. It can publish ReactionEvents and contains methods for adding and removing ReactionEventListeners as well as for firing reaction events and resetting the output modality. It contains a collection of all registered reaction event listeners to which reaction events can be fired.

The ReactionEventListener is the listener interface for subscribers who want to receive reaction events of an output modality. The class that is interested in processing reaction events implements this interface, and the object created with that class is registered with an output modality, using it's addReactionEventListener method. When a reaction event occurs, that object's onReactionEvent method is invoked.

The ReactionEvent is the event class which indicates that a reaction event occurred. It contains a status which is an enumeration and can take on the values started, completed and aborted. The reaction event also references the corresponding interaction of the document model which is presented by this reaction. It also contains a reference to the corresponding output modality which is the source and has fired this reaction event.

The EarconOutputModality is a subclass which implements the Earcon Output Modality by extending the abstract OutputModality base superclass. The Earcon Output Modality presents the interactions of the document models using earcons, which are non-verbal audio messages that are used in the computer user interface to provide information to the user about some computer object, operation or interaction (Blattner, Sumikawa & Greenberg, 1989). Each interaction is presented by a single-pitch inherited elementary earcon inheriting the timbre of the element type, pitch of the element level, spatial location of the element position, register of the operation and dynamics of the status. These single-pitch inherited elementary earcons are rendered simultaneously as parallel compound earcons (Brewster, Wright & Edwards, 1993) for all interactions that are going on in the user interface at the same time to reduce the length of time a compound audio message takes. The Earcon Output Modality has been introduced in detail in Chapter 5.2.2.

The TactonOutputModality is an alternative subclass which implements the Tacton Output Modality by extending the abstract OutputModality base superclass. The Tacton Output Modality presents the interactions of the document models using tactons or tactile icons. These are structured, abstract vibrotactile messages that can be used for presenting multidimensional information non-visually (Brewster & Brown, 2004). Each interaction is presented by a single-motive inherited elementary tacton inheriting the rhythm of the element type, tempo of the element level and intensity of the operation. These single-motive inherited elementary tactons are rendered sequentially as serial compound tactons for all interactions that are going on in the user interface at the same time. The Tacton Output Modality has been introduced in detail in Chapter 5.2.3.

The SpeechOutputModality is a third subclass which implements the Speech Output Modality by extending the abstract OutputModality base superclass. The Speech Output Modality presents the interactions of the document models using synthetic speech. Each interaction is presented by a inherited elementary utterance, inheriting the voice of the element type, pitch of the node level, spatial location of the node position, register of the operation and text of the content. These inherited elementary utterances are rendered sequentially as serial compound utterances for all interactions that are going on in the user interface at the same time. The Speech Output Modality has been introduced in detail in Chapter 5.2.4.

Automated Structured Observation

The Structured Observation package implements an automated structured observation research method. According to Bryman (2016), structured observation is a technique in which explicitly formulated rules for the observation and recording of behaviour are employed. The rules inform observers about what they should look for and how they should record behaviour. Each person who is part of the research is observed for a predetermined period of time using the same rules. These rules are articulated in what is usually referred to as an observation schedule. The automated structured observation research method has been introduced in detail in the previous Chapter 3. It contains sub packages for all observers, recorders, participants and observation schedules provided by the automated structured observation research method.

Observer

The Observer package contains all interfaces and classes of observers. An observer applies the observation schedule to a research subject. It presents the title, instructions and task of an exercise to the research subject. Afterwards it observes the research subject while he or she is solving the task. If the task has been successfully solved, the congratulations and the continuation is presented. Figure C.5 gives an overview over the Observer package:

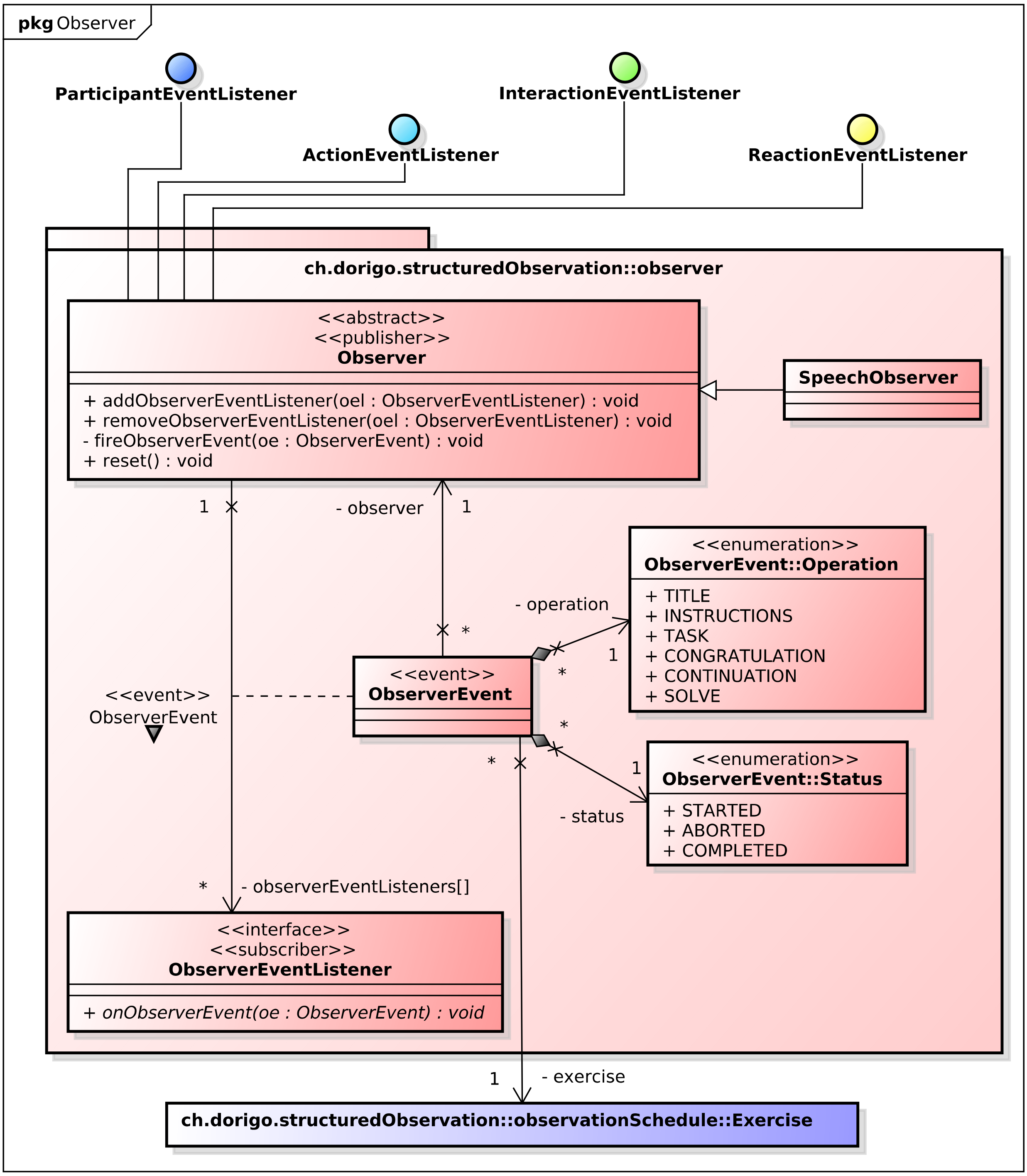


Figure C.5: Package Observer Overview

The Observer is the abstract base class for all observers. It implements the ParticipantEventListener, ActionEventListener, InteractionEventListener and

ReactionEventListener interfaces of the corresponding packages, so that it can receive and react on ParticipantEvents, ActionEvents, InteractionEvents and ReactionEvents provided by the different components. It publishes ObserverEvents and contains methods for adding and removing ObserverEventListeners as well as for firing observer events and resetting the observer. It contains a collection of all registered observer event listeners to which observer events can be fired.

The ObserverEventListener is the listener interface for subscribers who want to receive observer events of an observer. The class that is interested in processing observer events implements this interface, and the object created with that class is registered with an observer, using it’s addObserverEventListener method. When an observer event occurs, that object's onObserverEvent method is invoked.

The ObserverEvent is the event class which indicates that an observer event occurred. It contains a operation which is an enumeration and can take on the values title, instructions, task, congratulation, continuation and solve. It also contains a status which is an enumeration too and can take on the values started, completed and aborted. The reaction event also references the corresponding exercise of the observation schedule on which this observer event is performed. It also contains a reference to the corresponding observer which is the source and has fired this observer event.

The SpeechObserver is a subclass which implements the Speech Observer by extending the abstract Observer base superclass. The Speech Observer presents the title, instructions, task, congratulations and continuation texts of an exercise to the research subject using synthetic speech. In addition an applause sound is played if a task has been successfully solved.

Recorder

The Recorder package contains all interfaces and classes of recorders. The aim of a recorder is to receive and record all events that occur in the user interface as well as in the automated structured observation during the observation period, serialise an storing or streaming them over the Internet for later processing and analysis of this recorded data. Figure C.6 gives an overview over the Recorder package:

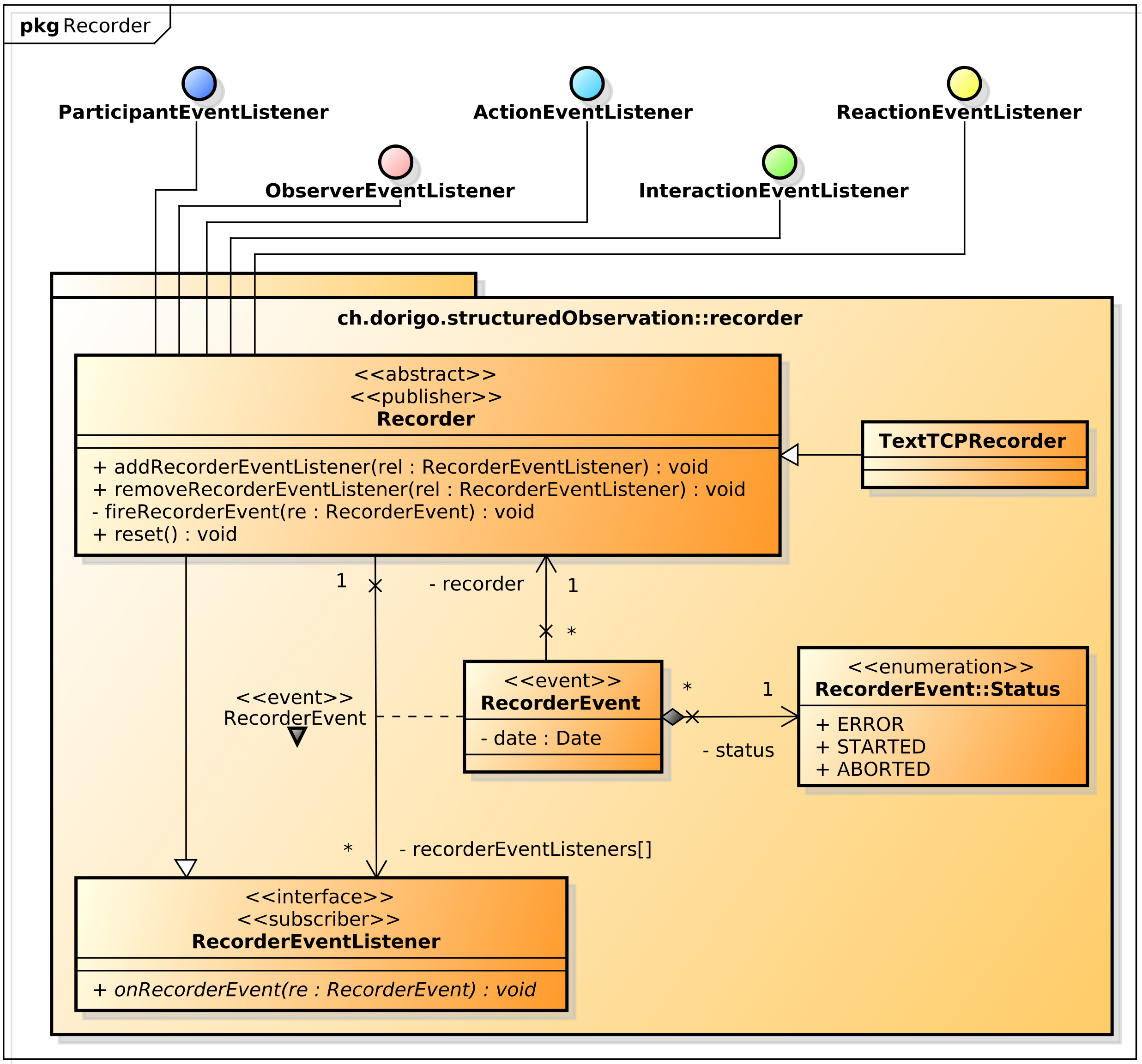


Figure C.6: Package Recorder Overview

The Recorder is the abstract base class for all recorders. It implements the ParticipantEventListener, ObserverEventListener, ActionEventListener, InteractionEventListener and ReactionEventListener interfaces of the corresponding packages, so that it can receive and record the ParticipantEvents, ObserverEvents, ActionEvents, InteractionEvents and ReactionEvents provided by the different components. It publishes RecorderEvents and contains all methods for adding and removing RecorderEventListeners as well as for firing recorder events and resetting the recorder. It contains a collection of all registered recorder event listeners to which recorder events can be fired.

The RecorderEventListener is the listener interface for subscribers who want to receive recorder events of a recorder. The class that is interested in processing recorder events implements this interface, and the object created with that class is registered with a recorder, using it’s addRecorderEventListener method. When a recorder event occurs, that object's onRecorderEvent method is invoked.

The RecorderEvent is the event class which indicates that a recorder event occurred. It contains the date at which the recorder event has occurred. The recorder event also contains a status which is an enumeration and can take on the values started, aborted and error. It also contains a reference to the corresponding recorder which is the source and has fired this recorder event.

The TextTCPRecorder is a subclass which implements the Text TCP Recorder by extending the abstract Recorder base superclass. Each received event is serialised to one single line of text. Each line starts with the time at which the event occurred in milliseconds since the start of the observation period, followed by the class name of the event and the name of the source class which has fired the event. Afterwoards the value of each attribute of the event is contained separated by tabulators. If a reference to an element node of the document or to an exercise of the observation schedule is serialised, the value of it’s “id” attribute is used. If a reference to a text node of the document is serialised, the first maximal 20 characters of it’s text content are used instead.

During the whole structured observation period, these recorded events are live streamed over the Internet to the Doky Server using a secured Transport Channel Protocol (TCP) connection. At the server, these events are stored to a file for each participant in a password protected directory for later data processing and data analysis. It is also possible to follow the ongoing structured observations live in real-time.

Participant

The Participant package contains all interfaces and classes of participants. According to Bryman (2016), a participant or research subject is a person who is taking part in the research and is observed for a predetermined period of time using the rules articulated in the observation schedule. Figure C.7 gives an overview over the Participant package:

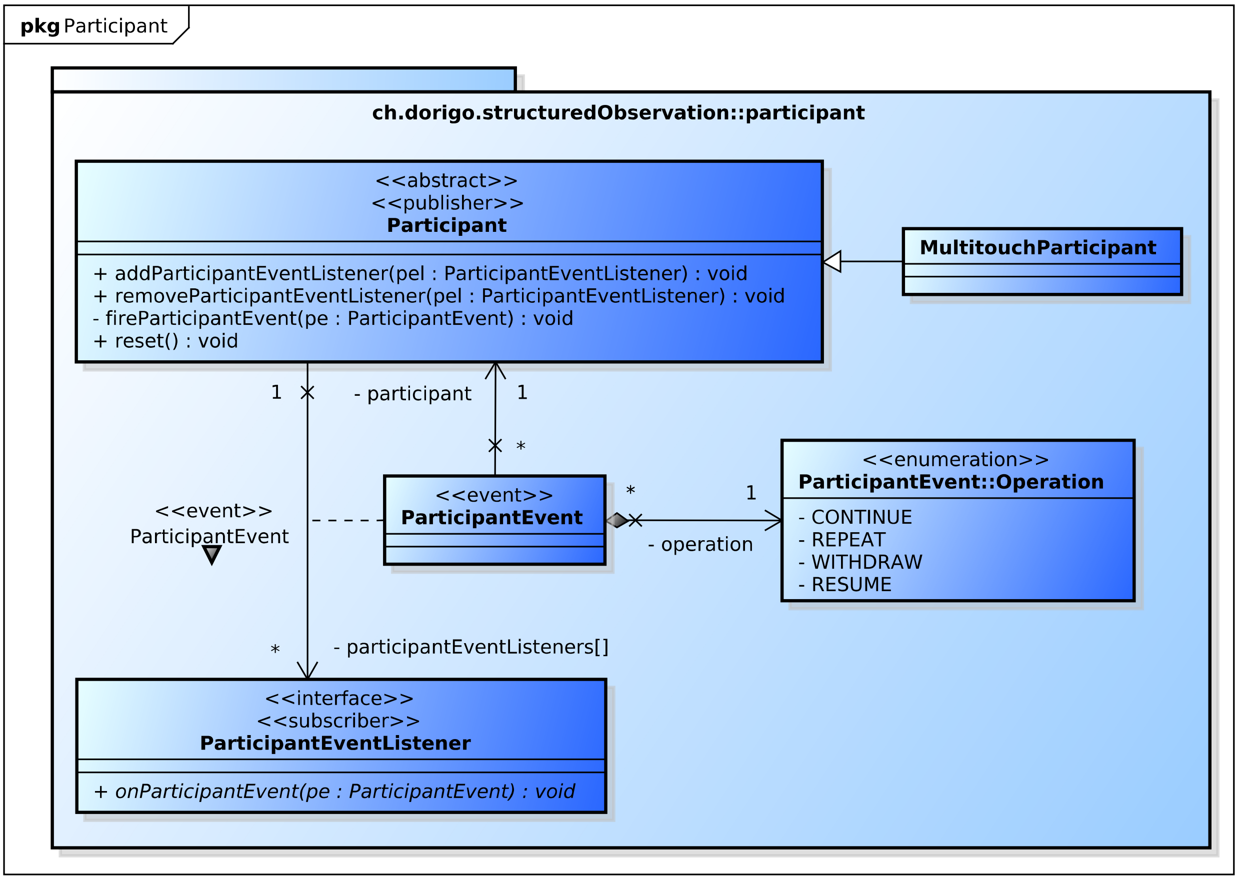


Figure C.7: Package Participant Overview

The Participant is the abstract base class for all participants. It publishes ParticipantEvents and contains methods for adding and removing ParticipantEventListeners as well as for firing participant events and resetting the participant. It contains a collection of all registered participant event listeners to which participant events can be fired.

The ParticipantEventListener is the listener interface for subscribers who want to receive participant events of a participant. The class that is interested in processing participant events implements this interface, and the object created with that class is registered with a participant, using it’s addParticipantEventListener method. When a participant event occurs, that object's onParticipantEvent method is invoked.

The ParticipantEvent is the event class which indicates that an action has been performed by a participant. It contains an operation which is an enumeration and can take on the values continue, repeat, withdraw and resume. It also contains a reference to the corresponding participant who is the source and has fired this participant event.

The MultitouchParticipant is a subclass which implements the Multitouch Participant by extending the abstract Participant base superclass. The Multitouch Participant allows the user to perform the actions provided by the observers by performing multi-touch gestures on the multi-touch screen or multi-touch pad of the mobile and wearable device. The research subject can listen to the instructions again at any time by double-tapping with one finger anywhere on the screen. He or she can continue with the next exercise by single-tapping with one finger anywhere on the screen. In addition the participant can withdraw at any time by pressing the home button. It is possible to continue at the same position at a later point in time by starting the Doky Structured Observation App again.

Observation Schedule

The Observation Schedule package contains all classes of an observation schedule. According to Bryman (2016), the aim of the observation schedule is to ensure that each participant's behaviour is systematically recorded so that it is possible to aggregate the behaviour of all those in the sample in respect of each type of behaviour being recorded. The rules that constitute the observation schedule should be as specific as possible in order to direct observers to exactly what aspects of behaviour they are supposed to be looking for. Each person who is part of the research is observed for a predetermined period of time using the same rules. Figure C.8 gives an overview over the Observation Schedule package:

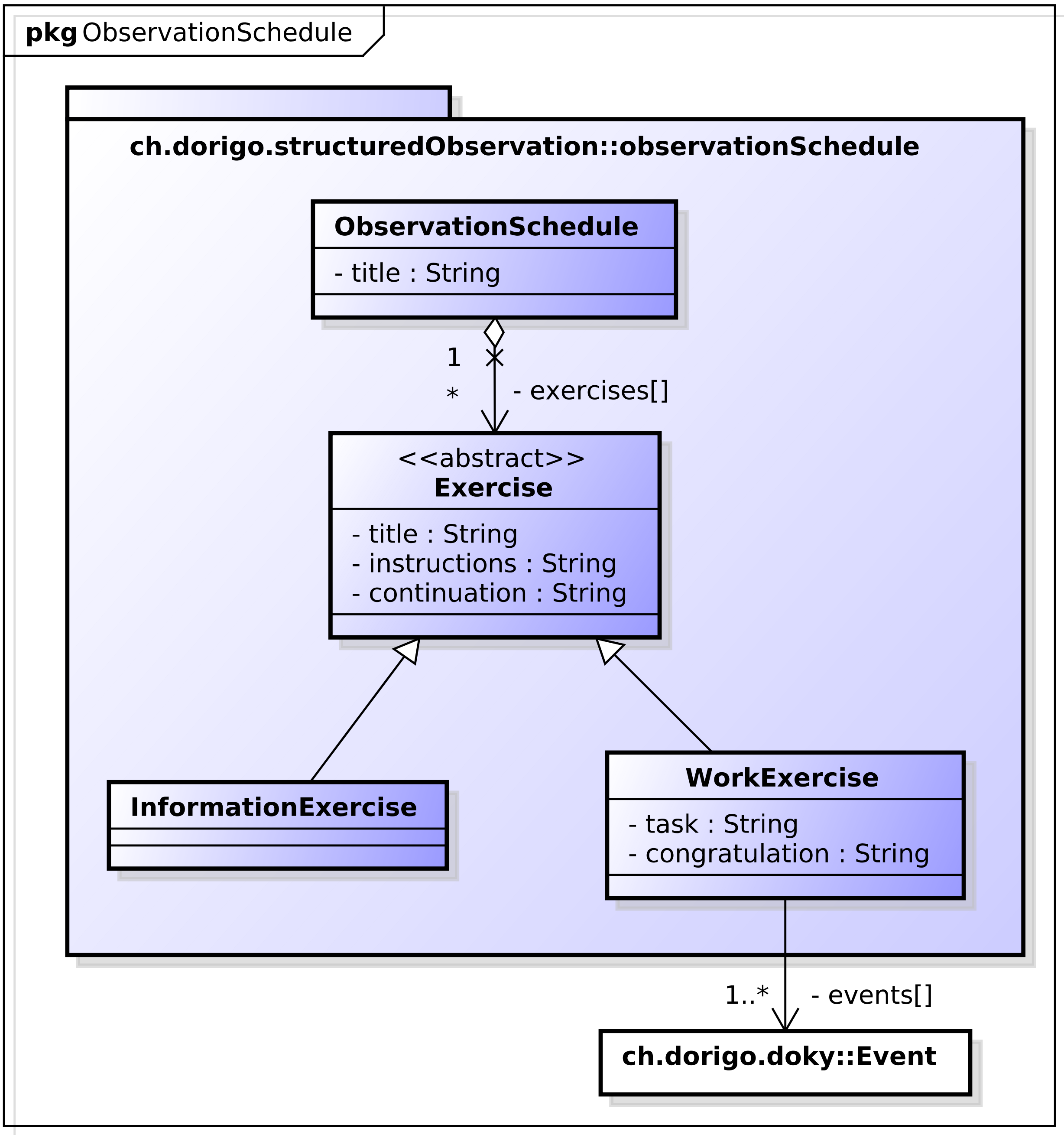


Figure C.8: Package Observation Schedule Overview

The ObservationSchedule class is the primary datatype for the entire observation schedule. Conceptually, it is the root of the observation schedule and provides the primary access to it's exercises. The observation schedule contains a title. In addition it consists of a collection of one or more exercises.

The Exercise is the abstract base class for all exercises. It contains a title, instructions as well as a continuation text. There are two subclasses of exercises: WorkExercises and InformationExercises.

A WorkExercise contains a task and a congratulations text. It also contains a collection of one or more Events formulating the goal of the task which have to occur in the given order in order for the task to be successfully solved.

An InformationExercise, on the other hand, is for information purposes only. It does not contain any task, congratulations text or events.

DOKY Structured Observation App

The DOKY Structured Observation App package contains the main class of the DOKY Structured Observation App. The DOKY Structured Observation App is an application of all the different components for conducting automated structured observations of research subjects performing exercises among a highly structured example document on their own mobile or wearable devices remotely over the Internet. Figure C.9 gives an overview over the DOKY Structured Observation App package:

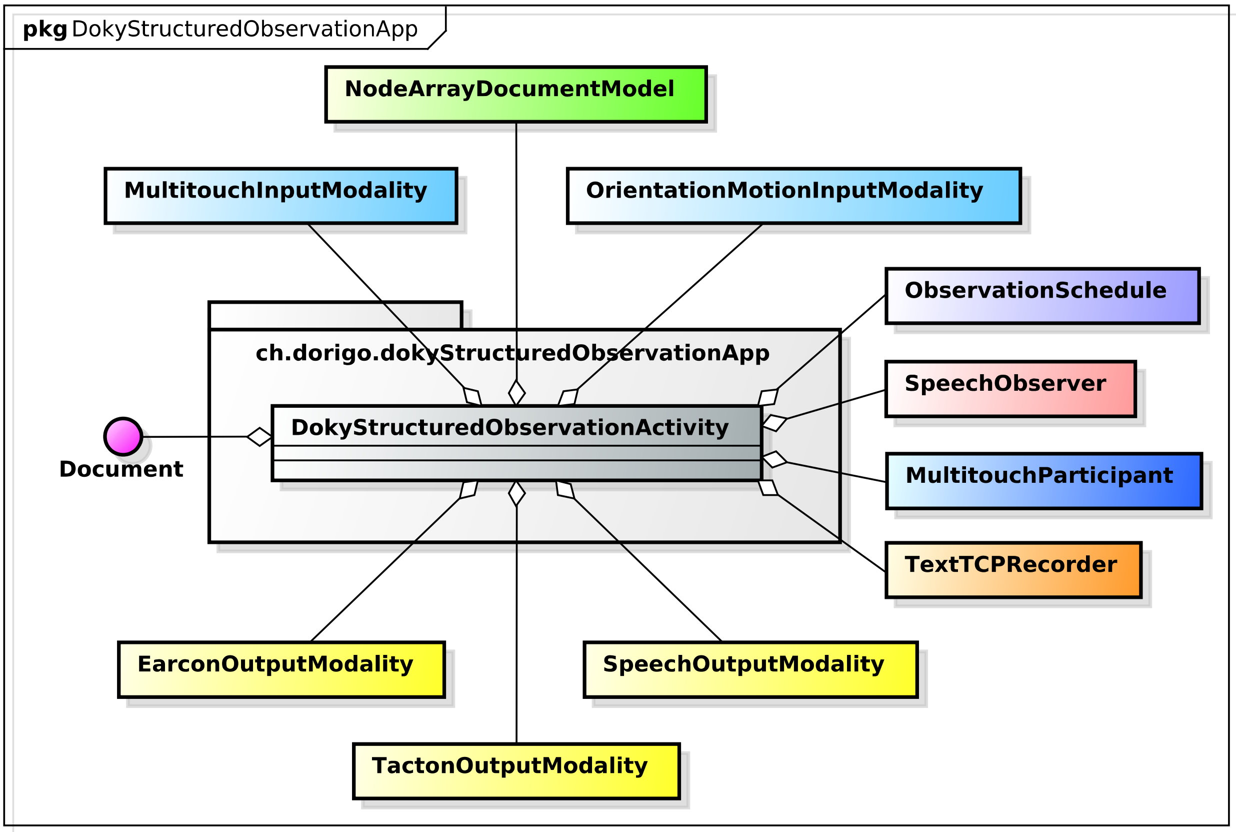


Figure C.9: Package DOKY Structured Observation App Overview

The DokyStructuredObservationActivity is the main class of the DOKY Structured Observation App. It parses the example document which is stored in a HTML file as Document Object Model and creates a Node Array Document Model of that document which is registered to receive action events of a Multitouch Input Modality and an Orientation Motion Input Modality object for reacting on the actions performed by the user. It also contains an Earcon Output Modality, Tacton Output Modality and Speech Output Modality object, which is registered to receive interaction events of the Node Array Document Model instance for presenting the interactions to the user.

The Doky Structured Observation Activity also parses the observation schedule which is stored in an XML file and creates an Observation Schedule object. It contains a Speech Observer instance which is registered to receive participant events of a Multitouch Participant object for reacting on the actions performed by the research subject. It also contains an instance of the Text TCP Recorder, which is registered to receive all events of the other component objects for recording and streaming these events over the Internet to the Doky Server.

After the Doky Structured Observation App has been started, for each exercise of the observation schedule, the instructions and than the task text are presented to the research subject by the speech observer. Afterwards the DOKY user interface with the example document is enabled and the research subject can start performing the task among the example document. When the research subject successfully completed the goal of the task, the user interface with the example document is disabled and the congratulations text is presented to the research subject by the speech observer. In addition an applause sound is played. After the congratulation the continuation text is presented to the research subject. While the research subject is solving a task the screen is black because the whole exercise is on a non-visual basis. After the whole observation schedule has been completed, the Doky Structured Observation App is terminated.

If the research subject would like to listen to the instructions again, the process of performing the task among the example document is interrupted and the user interface with the example document disabled and the instructions as well as the task text is presented to the research subject by the speech observer. Afterwards the user interface with the example document is enabled again and the process of performing the task among the example document continues. If the participant has withdrawn at a time, the Doky Structured Observation App will be terminated immediately and the current status of the structured observation is saved. It is possible to continue at the same position at a later point in time by starting the Doky Structured Observation App again.

Document Object Model (DOM)

The Document Object Model package contains all interfaces of the Document Object Model (DOM). The Document Object Model (DOM), developed by Van Kesteren, Aryeh, Russell & Berjon (2015), is a platform- and language-neutral application programming interface (API) that allows programs and scripts to dynamically access and update the content, structure and style of structured documents. The Document Object Model (DOM) has been introduced in detail in Chapter 2.4.1. Figure C.10 gives an overview over the Document Object Model package:

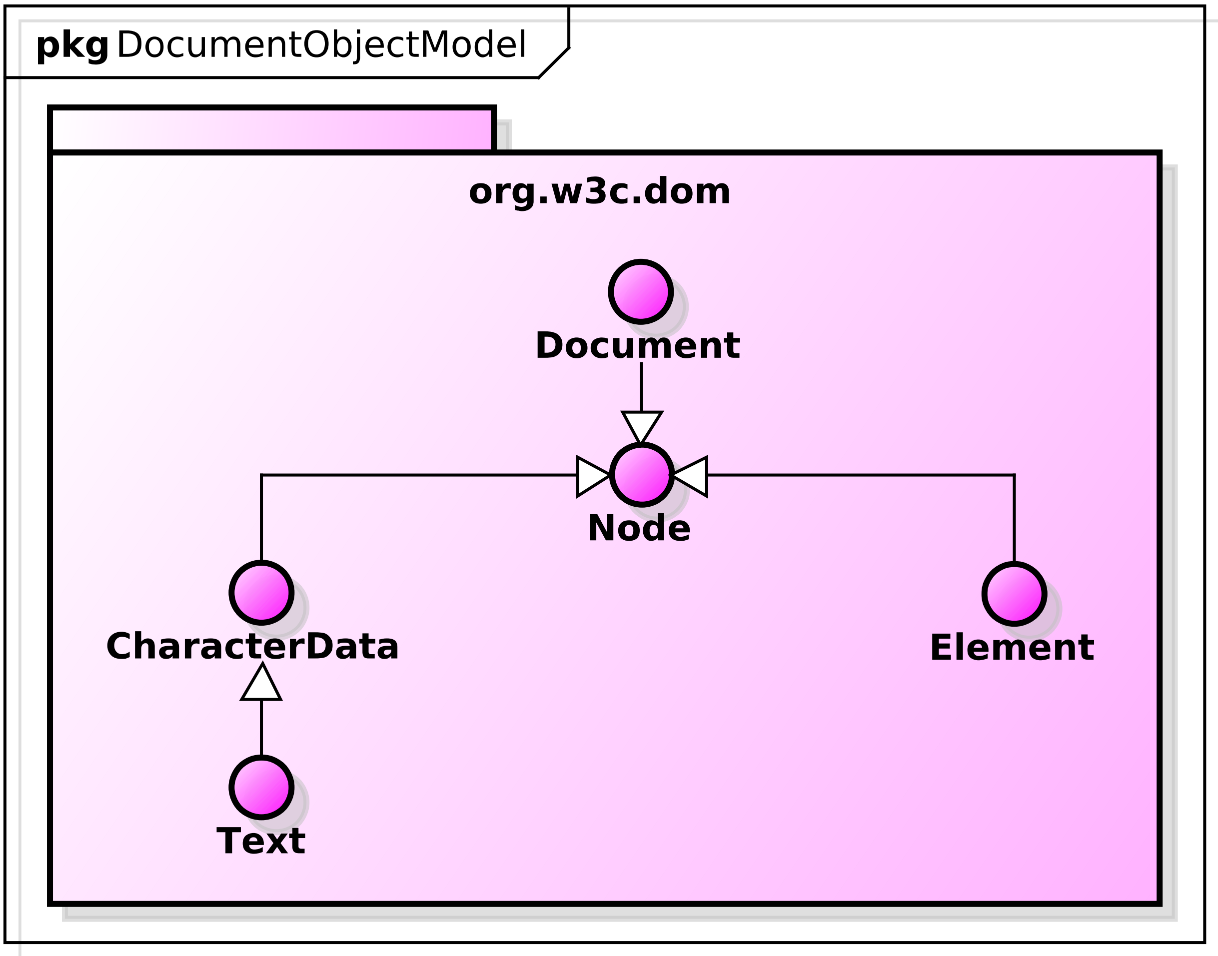


Figure C.10: Package Document Object Model Overview

In the DOM, documents have a logical structure which is very much like a tree. The DOM presents documents as a hierarchy of Node objects that also implement other, more specialized interfaces. Some types of nodes may have child nodes of various types, and others are leaf nodes that cannot have anything below them in the document structure. A child is an immediate descendant node of a node.

The Node interface is the primary datatype for the entire Document Object Model. It represents a single node in the document tree. The most important subtypes of nodes are document, element and text nodes.

The Document node represents the entire document. Conceptually, it is the root of the document tree, and provides the primary access to the document's data. The root node is a node that is not a child of any other node. All other nodes are children or other descendants of the root node. Element and Text nodes cannot exist outside the context of a Document node.

Each document contains one or more elements. The Element node represents an element in a document. Each element has a type, identified by name, and may have a set of attributes. Each attribute has a name and a value.

The Text interface inherits from CharacterData and represents the textual content of an Element or attribute. If there are no other elements inside an element's content, the text is contained in a single object implementing the Text interface that is the only child of the element. If there are other elements contained, these element and text node items form a list of children of the element. Two nodes are siblings if they have the same parent node.

Conclusions

A flexible platform-independent and event-driven software architecture implementing the DOKY user interface as well as the automated structured observation research method employed for the evaluation of the effectiveness of the proposed user interface has been presented. Because it is platform- and language-neutral, it can be used in a wide variety of platforms, environments and applications for mobile and wearable devices. Each component is defined by interfaces and abstract classes only, so that it can be easily changed or extended, and grouped in a semantically self-containing package.

The Doky package implements the DOKY user interface, which is a multi-model user interface for non-visual presentation, navigation and manipulation of structured documents on mobile and wearable devices. It contains sub-packages for all document models, input modalities and output modalities provided by the DOKY user interface.

The Document Models package contains all interfaces and classes of modality-neutral document models for presentation, navigation and manipulation of structured documents. A document model is a collection of descriptions of data structures and their contained fields, together with the operations or functions that manipulate them. The Node Array Document Model is an alternative modality-neutral document model for presentation, navigation and manipulation of structured documents. In contrast to the tree-like document representation employed in the Document Object Model (DOM), the element and text nodes of the document are organised in a two-dimensional array.

The Input Modalities package contains all interfaces and classes of input modalities. An input modality enables the user to perform the actions provided by the document models by stimulating s specific hardware human input device (HID). The Multitouch Input Modality allows the user to perform the actions provided by the document models by performing multi-touch gestures on the multi-touch screen or multi-touch pad of the mobile and wearable device. The Orientation Motion Input Modality allows the user to perform the actions provided by the document models by performing motion gestures by moving the mobile or wearable device itself.

The Output Modalities package contains all interfaces and classes of output modalities. An output modality presents the interactions of the document models to the user in a modality specific way by stimulation of one or more human senses using specific hardware human output devices. The Earcon Output Modality presents the interactions of the document models using earcons, which are non-verbal audio messages that are used in the computer user interface to provide information to the user about some computer object, operation or interaction. The Tacton Output Modality presents the interactions of the document models using tactons or tactile icons. These are structured, abstract vibrotactile messages that can be used for presenting multidimensional information non-visually. The Speech Output Modality presents the interactions of the document models using synthetic speech.

The Structured Observation package implements the automated structured observation research method. Structured observation is a technique in which explicitly formulated rules for the observation and recording of behaviour are employed. The rules inform observers about what they should look for and how they should record behaviour. Each person who is part of the research is observed for a predetermined period of time using the same rules. These rules are articulated in what is usually referred to as an observation schedule. It contains sub packages for all observers, recorders, participants and observation schedules provided by the automated structured observation research method.

The Observer package contains all interfaces and classes of observers. An observer applies the observation schedule to a research subject. It presents the title, instructions and task of an exercise to the research subject. Afterwards it observes the research subject while he or she is solving the task. If the task has been successfully solved, the congratulations and the continuation is presented. The Speech Observer presents the title, instructions, task, congratulations and continuation texts of an exercise to the research subject using synthetic speech. In addition an applause sound is played if a task has been successfully solved.

The Recorder package contains all interfaces and classes of recorders. The aim of a recorder is to receive and record all events that occur in the user interface as well as in the automated structured observation process during the observation period, serialise an storing or streaming them over the Internet for later processing and analysis of this recorded data. The Text TCP Recorder serialised each received event in one single line of text. These recorded events are live streamed over the Internet to the Doky Server using a secured Transport Channel Protocol (TCP) connection.

The Participant package contains all interfaces and classes of participants. A participant or research subject is a person who is taking part in the research and is observed for a predetermined period of time using the rules articulated in the observation schedule. The Multitouch Participant allows the user to perform the actions provided by the observers by performing multi-touch gestures on the multi-touch screen or multi-touch pad of the mobile and wearable device.

The Observation Schedule package contains all classes of an observation schedule. The aim of the observation schedule is to ensure that each participant's behaviour is systematically recorded so that it is possible to aggregate the behaviour of all those in the sample in respect of each type of behaviour being recorded. The rules that constitute the observation schedule should be as specific as possible in order to direct observers to exactly what aspects of behaviour they are supposed to be looking for. Each person who is part of the research is observed for a predetermined period of time using the same rules.

The Doky Structured Observation App package contains the main class of the Doky Structured Observation App. The Doky Structured Observation App is an application of all the different components for conducting automated structured observations of research subjects performing exercises among a highly structured example document on their own mobile or wearable devices remotely over the Internet.

The Document Object Model package contains all interfaces of the Document Object Model (DOM). The Document Object Model (DOM) is a platform- and language-neutral application programming interface (API) that allows programs and scripts to dynamically access and update the content, structure and style of structured documents.

DOKY